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

DENIS CHETVERIKOV
UCLA, Los Angeles, CA 90095, U.S.A.

BRADLEY LARSEN
Stanford University, Stanford, CA 94305, U.S.A. and NBER

CHRISTOPHER PALMER
Haas School of Business, University of California at Berkeley, Berkeley, CA 94720, U.S.A.
IV QUANTILE REGRESSION FOR GROUP-LEVEL TREATMENTS, WITH AN APPLICATION TO THE DISTRIBUTIONAL EFFECTS OF TRADE

BY DENIS CHETVERIKOV, BRADLEY LARSEN, AND CHRISTOPHER PALMER

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. Our approach applies to a broad range of settings including labor, public finance, industrial organization, urban economics, and development; we illustrate its usefulness with several such examples. Finally, an empirical application of our estimator finds that low-wage earners in the United States from 1990 to 2007 were significantly more affected by increased Chinese import competition than high-wage earners.

KEYWORDS: Quantile regression, instrumental variables, panel data, income inequality, import competition.

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 thank the editor and four anonymous referees for their detailed comments and feedback; Moshe Buchinsky, Ivan Canay, Brigham Frandsen, Antonio Galvao, Wenshu Guo, Jerry Hausman, Rosa Matzkin, Whitney Newey, and Christopher Taber for helpful comments; and Sam Hughes, Yuqi Song, and Caio Waisman for meticulous research assistance. We are especially grateful to Jin Hahn for many useful discussions.
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.). 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.

In Section 3, we introduce our estimation approach, which we refer to as grouped IV quantile regression. The estimator is computationally simple to implement and consists of two steps: (i) 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 (ii) 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. Section 3 also discusses Monte Carlo simulations (found in Appendix A of the Supplemental Material (Chetverikov, Larsen, and Palmer (2016))) 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.

Section 4 provides a variety of examples illustrating the use of the grouped IV quantile regression estimator. In particular, we use examples from Angrist and Lang (2004), Larsen (2014), Palmer (2011), and Backus (2014) to illustrate applicability of our estimator. In addition to these examples, the grouped

---

2Similar terminology was used, for example, by Altonji and Matzkin (2005).  
3Even 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)).
quantile approach can apply to a wide range of settings in labor, industrial organization, trade, public finance, development, and other applied fields.

We derive theoretical properties of the estimator in Section 5. 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 heteroscedasticity-robust variance estimators for 2SLS, making inference particularly simple. In the Supplemental Material, we also discuss clustered standard errors, and we show how to construct confidence bands for the coefficient of interest which hold uniformly over a set of quantiles via multiplier bootstrap procedure.

Section 6 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.

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 $L^\infty(U)$ represents the set of bounded functions on $U$. With some abuse of notation, $\ell^\infty(U)$ also denotes the set of component-wise bounded vector-valued functions on $U$. 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 Supplemental Material.

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 a simple version of the model, which we consider as most appealing in empirical work, and then present the general version of our model, which allows for more flexible distributional effects. Our estimator and theoretical results apply to both the general and simple versions of the model.

In the simple version of the model, we assume that the $u$th quantile of the conditional distribution of $y_{ig}$ is given by

$$Q_{y_{ig}|\tilde{z}_{ig}, x_{g}, \varepsilon_{g}}(u) = \tilde{z}_{ig}' \gamma(u) + x_{g}' \beta(u) + \varepsilon_{g}(u), \quad u \in U,$$

where $Q_{y_{ig}|\tilde{z}_{ig}, x_{g}, \varepsilon_{g}}(u)$ is the $u$th conditional quantile of $y_{ig}$ given $(\tilde{z}_{ig}, x_{g}, \varepsilon_{g})$, $\tilde{z}_{ig}$ is a $(d_z - 1)$-vector of observable individual-level covariates (which we sometimes refer to as micro-level covariates), $x_{g}$ is a $d_x$-vector of observable group-level covariates ($x_{g}$ contains a constant), $\gamma(u)$ and $\beta(u)$ are $(d_z - 1)$- and $d_x$-vectors of coefficients, $\varepsilon_{g} = \{\varepsilon_{g}(u), u \in U\}$ is a set of unobservable group-level random scalar shifters, and $U$ is a set of quantile indices of interest. Here, $\gamma(u)$ and $\beta(u)$ represent the effects of individual- and group-level covariates, respectively. In this paper, we are primarily interested in estimating $\beta(u)$, although we also provide some new results on estimating $\gamma(u)$.

In the more general version of the model, of which (1) is a special case, we assume that the $u$th quantile of the conditional distribution of $y_{ig}$ is given by

$$Q_{y_{ig}|\tilde{z}_{ig}, x_{g}, \alpha_{g}}(u) = \tilde{z}_{ig}' \alpha_{g}(u), \quad u \in U,$$

$$\alpha_{g,1}(u) = x_{g}' \beta(u) + \varepsilon_{g}(u), \quad u \in U,$$

One interpretation of the term $\varepsilon_{g}(u)$ in (1) is that it accounts for all unobservable group-level covariates $\eta_{g}$ that affect the distribution of $y_{ig}$ 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.
where \( Q_{yig|z_{ig}, x_g, \alpha_g}(u) \) is the \( u \)th conditional quantile of \( y_{ig} \) given \((z_{ig}, x_g, \alpha_g)\), \( z_{ig} \) is a \( d_z \)-vector of observable individual-level covariates, \( \alpha_g = \{\alpha_g(u), u \in \mathcal{U}\} \) is a set of (random) group-specific effects with \( \alpha_{g,1}(u) \) being the first component of the vector \( \alpha_g(u) = (\alpha_{g,1}(u), \ldots, \alpha_{g,d_g}(u))^\prime \), and all other notation is the same as above. In this model, we assume that the response variable \( y_{ig} \) satisfies the quantile regression model in (2) 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 (3), in which we are interested in estimating the vector of coefficients \( \beta(u) \).

Observe that the model (1) is a special case of the model (2)–(3). Indeed, setting \( z_{ig} = (1, z_{ig})^\prime \) and assuming that \( (\alpha_{g,2}(u), \ldots, \alpha_{g,d_g}(u))^\prime = \gamma(u) \) for some non-stochastic \((d_z - 1)\)-vector \( \gamma(u) \) and all \( g = 1, \ldots, G \) in the model (2)–(3) gives the model (1) after substituting (3) into (2). The model (2)–(3) is more general, however, because it allows all coefficients of individual-level covariates to vary across groups via group-specific effects \( \alpha_g(u) \), and it also allows to study not only location shift effects of the group-level covariates \( x_g \) but also their interaction effects. Therefore, throughout this paper, we study the model (2)–(3).

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 (1) 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 (2)–(3) 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 (2) would capture this idea by allowing the researcher to specify a linear regression model of the form of (3) for the component of \( \alpha_g \) 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.\(^5\)

In many applications, it is likely that the group-level covariates \( x_g \) may be endogenous in the sense that \( E[x_g \epsilon_g(u)] \neq 0 \), 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

\(^5\)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 (3) for each effect.
\[ E[w_g \varepsilon_g(u)] = 0 \text{ for all } u \in U, \ E[w_g x_g'] \text{ is nonsingular, and } y_{ig} \text{ is independent of } w_g \text{ conditional on } (z_{ig}, x_g, \alpha_g). \]

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 \( \alpha_g \). It implies, in particular, that \( Q_{y_{ig}|z_{ig}, x_g, \alpha_g, w_g}(u) = z'_{ig} \alpha_g(u) \) for all \( u \in U \).

We assume that a researcher has data on \( G \) groups and \( N_g \) individuals within group \( g = 1, \ldots, G \). Thus, the data consist of observations on \( (z_{ig}, y_{ig}), i = 1, \ldots, N_g, x_g, \) and \( w_g \) for \( g = 1, \ldots, G \). Throughout the paper, we denote \( N_G = \min_{1 \leq g \leq G} N_g \). For our asymptotic theory in Section 5, 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 \( G^{3/2}(\log N_G)/N_G \to 0 \) 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 \( E[z_{ig,k} \varepsilon_g(u)] = 0 \text{ for all } u \in U \), 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 the applicability of our results.

\(^6\)The assumption that \( E[w_g \varepsilon_g(u)] = 0 \) holds jointly for all \( u \in U \) should not be confused with requiring quantile crossing. To understand it, assume, for example, that \( \varepsilon_g(u) = \varepsilon(u, \eta_g) \) where \( \eta_g \) is a vector of group-level omitted variables in regression (3). 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 \( E[\varepsilon(u, \eta_g)|w_g] = 0 \) does not depend on \( w_g \), which occurs (for example) if \( \eta_g \) 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 \( x_g \).

\(^7\)The setting we model differs from other IV quantile settings, such as Chernozhukov and Hansen (2005, 2006, 2008). Consider, for simplicity, our model (1) and assume that \( U = [0, 1] \). Then the Skorohod representation implies that \( y_{ig} = \tilde{z}_{ig} \gamma(u_{ig}) + x'_g \beta(u_{ig}) + \varepsilon_g(u_{ig}) \) where \( u_{ig} \) is a random variable that is distributed uniformly on \( [0, 1] \) and is independent of \( (\tilde{z}_{ig}, x_g, \varepsilon_g) \). Here, one can think of \( u_{ig} \) as an 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 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.
Our problem in this paper is different from that studied in Koenker (2004), Kato, Galvao, and Montes-Rojas (2012), and Kato and Galvao (2011). Specifically, they considered the panel data quantile regression model

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

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 \( \{\alpha_g(u), u = 1, \ldots, G\} \) 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 (1) by setting \( \alpha_g(u) = x_g'\beta(u) + \varepsilon_g(u) \)). They showed that their estimator is asymptotically zero-mean normal if \( G^2(\log G)/N_G \to 0 \) as \( G \to \infty \). Making further progress, Kato and Galvao (2011) suggested an interesting smoothed quantile regression estimator of \( \gamma(u) \) that is asymptotically zero-mean normal if \( G/N_G \to 0 \). 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 nonlinear panel data models. Formally, they considered a nonlinear panel data model defined by the following equation:

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

8Our 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 multidimensional) 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 \( \alpha_g(\cdot) = \alpha_g(\cdot) \) to depend on \( i \).

9To 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 \( G^2(\log G)/N_G \to 0 \), assume, for simplicity, that \( d_1 = 1, d_2 = 2, x_g \) and the second component of \( z_{ig} \) are constants, that is, \( x_g = 1 \) and \( z_{ig} = (\tilde{z}_{ig}', 1)' \). Then our model (2)–(3) reduces to \( Q_{y_{ig}|z_{ig},\varepsilon_g,\alpha_g(u)} = \tilde{z}_{ig}(\beta(u) + \varepsilon_g(u)) + \alpha_g(u) \), 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 5. 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.
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 $\beta$) without assuming that $\varepsilon_g$ 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 $G^{2/3} (\log N_G)/N_G \rightarrow 0$ as $G \rightarrow \infty$, with the improvement coming from a better control of the residuals in the Bahadur representation.

### 3. ESTIMATOR

In this section, we develop our estimator, which we refer as grouped IV quantile regression. 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 $U$ of indices of interest, estimate $u$th quantile regression of $y_{ig}$ on $z_{ig}$ using the data \{$(y_{ig}, z_{ig}) : i = 1, \ldots, N_g$\} by the classical quantile regression estimator of Koenker and Bassett (1978):

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

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

**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,

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

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

---

10 The use of a 2SLS regression on the second stage of our estimator is dictated by our assumption that $\varepsilon_g(u)$ is (mean)-uncorrelated with $w_g$: $E[w_g \varepsilon_g(u)] = 0$. If, instead, we assumed that $\varepsilon_g(u)$ is median-uncorrelated with $w_g$, a concept developed in Komarova, Severini, and Tamer (2012), the second stage of our estimator would be an IV quantile regression developed in Chernozhukov and Hansen (2006). In this case, our method would be a quantile-after-quantile estimator.
Several special cases of our estimator are worth noting. First, when the model is given by equation (1), the steps of our estimator consist of (i) 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 (ii) 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 \( u \)th quantile of the outcome variable \( y_{ig} \) within each group. Third, if \( x_{g} \) is exogenous, that is, \( \mathbb{E}[x_{g}\varepsilon_{g}(u)] = 0 \), 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 (1), 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 data set 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 data sets with many fixed effects. Second, even if no group-level unobservables exist (consider model (1) 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) 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.

Monte Carlo simulations in Appendix A of the Supplemental Material highlight the performance of our estimator for \( \beta(u) \) in (1) relative to the traditional Koenker and Bassett (1978) estimator (which ignores endogeneity of \( x_{g} \)).

---

11In Monte Carlo experiments in Appendix A of the Supplemental Material, 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.

12One 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 Section 4.
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 below, standard errors for our estimator $\hat{\beta}(u)$ may be obtained using standard heteroscedasticity-robust (Section 5) or clustering (Appendix E of the Supplemental Material) approaches for 2SLS or OLS as if there were no first stage. 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.

4. 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. 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.13

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

$$Q_{y_{igt}|x_{igt}}(0.2) = \alpha_g(0.2) + \beta_j(0.2) + \gamma_t(0.2) + \delta(0.2)m_{igt} + \lambda(0.2)s_{igt} + \xi_{igt}(0.2),$$

where the left-hand side represents the second decile of student test scores within a group, $x_{igt} = (m_{igt}, s_{igt}, \xi_{igt}, \alpha_g, \beta_j, \gamma_t)$, and a group is a grade $g \times$ school $j \times$ year $t$ cell. The variables $s_{igt}$ and $m_{igt}$ denote the class size and the

---

13This is in contrast to settings where the endogeneity exists at the individual level, that is, when the individual unobserved heterogeneity is correlated with treatment. Such situations require a different approach than the one presented here, for example, Chernozhukov and Hansen (2005), Abadie, Angrist, and Imbens (2002), or the other approaches referenced in Section 1.
fraction of Metco students within each $g \times j \times t$ cell, and $\alpha_g$, $\beta_j$, and $\gamma_t$ represent grade, school, and year effects, respectively. The unobserved component $\xi_{gjt}$ is analogous to $\varepsilon(0.2)$ in our model (1).

**Angrist and Lang (2004)** estimated equation (5) by OLS, which is equivalent to the non-IV application of our estimator with no micro-level covariates. Similarly 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 (5) by 2SLS, they were unable to address the causal *distributional* effects of Metco exposure.

In estimating (5), **Angrist and Lang (2004)** used heteroscedasticity-robust standard errors, which we demonstrate in Section 5 is valid. The extension in Appendix E of the Supplemental Material implies that the authors could have instead allowed for clustering across groups in computing standard errors (e.g., clustering at the school level given a sufficient number of schools).

**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. Similarly 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 a state-year combination $(s, t)$, and micro-level data consist of teachers within a particular state in a given year. The conditional $u$th quantile of teacher quality among teachers who began teaching in state $s$ in year $t$ is modeled as

$$Q_{qist|Law_{st},z_{ist}}(u) = \gamma_s(u) + \lambda_t(u) + Law_{st}' \delta(u) + \varepsilon_{st}(u),$$

where $Law_{st}$ is a vector of dummies capturing the type of certification tests required for licensure in state $s$ and year $t$, $\gamma_s(u)$ and $\lambda_t(u)$ are state and year effects, 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 $u$th quantile of quality in a given state-year cell. The second stage is obtained via OLS. **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 (i) estimating quantile regression of $q_{ist}$ on a vector $z_{ist}$ (which would include a measure of the percent...
minority students) separately for each \((s, t)\) group and saving each group-level estimate for the coefficient corresponding to the percent minority variable; and (ii) 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 can drastically increase 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.\(^{14}\)

**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.\(^{15}\)

The model is

\[
\Delta Q_{y_{igt}}|x_g, e_g, \epsilon_g(u) = \beta(u) \cdot \text{suburbanization}_g + x'_g \gamma_1(u) + e_g(u),
\]

\[
\text{suburbanization}_g = \pi(u) \cdot \text{planned highway rays}_g + x'_g \gamma_2(u) + v_g(u),
\]

where \(\Delta Q_{y_{igt}}|x_g, e_g, \epsilon_g(u)\) is the change in the \(u\)th 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 \(\text{planned highway rays}_g\) is uncorrelated with \(e_g(u)\) and \(v_g(u)\). The variable \(\text{suburbanization}_g\) 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

\(^{14}\)Note also that 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 estimating (6) using the grouped approach was significantly faster than estimating (6) in a single standard quantile regression. See also Appendix A of the Supplemental Material for further discussion of computational advantages of the grouped quantile approach.

\(^{15}\)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.
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 $\rho_{imt}$ 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 $\rho_{imt}$ satisfies the following quantile regression model:

$$Q_{\rho_{imt}|_{cmt, n_{mt}, e_{mt}(u)}}(u) = \beta_t(u) + c_{mt}\beta_c(u) + g(n_{mt}, u) + e_{mt}(u),$$

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 $e_{mt}$ 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 (7) represents an application of our estimator when group-level 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. Backus (2014) reported standard errors clustered at the market level, which we demonstrate are valid in Appendix E of the Supplemental Material.

5. **ASYMPTOTIC THEORY**

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

5.1. **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 $N_G = \min_{s=1,\ldots,G} N_g$. We start with specifying our main assumptions.
ASSUMPTION 1—Design: (i) Observations are independent across groups. (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 \((x_g, \alpha_g)\).

ASSUMPTION 2—Instruments: (i) For all \( u \in \mathcal{U} \) and \( g = 1, \ldots, G \), \( E[w_g \varepsilon_g(u)] = 0 \). (ii) As \( G \to \infty \), \( G^{-1} \sum_{g=1}^{G} E[x_g w'_g] \to Q_{xw} \) and \( G^{-1} \times \sum_{g=1}^{G} E[w'_g w_g] \to Q_{ww} \) where \( Q_{xw} \) and \( Q_{ww} \) are matrices with 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 \( w_k \) conditional on \((z_{ig}, x_g, \alpha_g)\). (iv) For all \( g = 1, \ldots, G \), \( E[\|w_g\|^{4+c_M}] \leq C_M \).

ASSUMPTION 3—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 (e.g., 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 of the Supplemental Material. Assumption 1(ii) allows for interdependence (clustering) within groups but imposes the restriction that the interdependence between observations within the group \( g \) is fully controlled for by the group-level covariates \( x_g \) and the group-specific effect \( \alpha_g \). Assumption 2 is our main identification condition. Note that Assumption 2 allows for internal instruments. In particular, if \( w_k = N_g^{-1/2} \sum_{i=1}^{N_g} z_{ig,k} \) 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.\(^{16}\)

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 \( y_{ig} \) given \((z_{ig}, x_g, \alpha_g)\) (dependence of \( f_{g}(\cdot) \) on \((z_{ig}, x_g, \alpha_g)\) is not shown explicitly for brevity of notation). Also denote \( B_{g}(u, c) = (z'_{ig} \alpha_g(u) - c, z'_{ig} \alpha_g(u) + c) \) for \( c > 0 \). We will assume the following regularity conditions:

ASSUMPTION 4—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 \( E_g[z_{ig} z'_{ig}] \) are bounded from below by \( c_M \).

\(^{16}\)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 \( N \to \infty \).
ASSUMPTION 5—Coefficients: For all \( u_1, u_2 \in \mathcal{U} \) and \( g = 1, \ldots, G \), \( \| \alpha_g(u_2) - \alpha_g(u_1) \| \leq C_L |u_2 - u_1| \).

ASSUMPTION 6—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 \( J : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^{d_x \times d_y} \), \( G^{-1} \sum_{g=1}^G E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_g u_g'] \to J(u_1, u_2) \) uniformly over \( u_1, u_2 \in \mathcal{U} \). (iii) For all \( u_1, u_2 \in \mathcal{U} \), \( |\varepsilon_g(u_2) - \varepsilon_g(u_1)| \leq C_L |u_2 - u_1| \).

ASSUMPTION 7—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 \( y \in B_g(u, c_f) \) and \( |f_g'(z' x_g \alpha_g(u))| \geq c_f \). (ii) For all \( u \in \mathcal{U} \) and \( g = 1, \ldots, G \), \( f_g(y) \leq C_f \) for all \( y \in B_g(u, c_f) \) and \( f_g(z' x_g \alpha_g(u)) \geq c_f \).

ASSUMPTION 8—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 \( z_g \) 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 \((w_g, \varepsilon_g)\) 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, e.g., Chernozhukov (2005) and Chernozhukov and Fernández-Val (2011)).

5.2. Results

We now present our main results. In Theorem 1, we derive the asymptotic distribution of our estimator. In Theorem 2, we show how to estimate the asymptotic covariance of our estimator. For brevity of the paper, further results are relegated to Appendices C–E of the Supplemental Material. In particular, in Appendix C, we describe a multiplier bootstrap method for constructing uniform over \( u \in \mathcal{U} \) confidence intervals for \( \beta(u) \) and prove its validity relying on results from Chernozhukov, Chetverikov, and Kato (2013). In Appendix D, we present an approach for uniform inference on \( \{ \alpha_{g,1}(u), g = 1, \ldots, G \} \) in
the model (2)–(3) by constructing the confidence bands \([\hat{\alpha}_{g,1}(u), \hat{\alpha}_{g,1}(u)]\) that cover the true group-specific effects \(\alpha_{g,1}(u)\) for all \(g = 1, \ldots, G\) simultaneously with probability approximately \(1 - \alpha\). In Appendix E, we consider clustered standard errors.

The first theorem derives the asymptotic distribution of our estimator.

**THEOREM 1—Asymptotic Distribution:** Let Assumptions 1–8 hold. Then

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

where \(\mathbb{G}(\cdot)\) is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function \(C(u_1, u_2) = SJ(u_1, u_2)S',\) where \(S = (Q_{wu}Q_{ww}^{-1}Q_{wu})^{-1}Q_{wu}Q_{ww}^{-1}, Q_{wu}\) 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 \(u \in U\). In addition, Theorem 1 implies that, for any \(u \in U,\)

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

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 \(u \in U,\) we have to deal with \(G\) independent quantile processes \([\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u), u \in U]\). Since \(G \to \infty,\) classical functional central limit theorems do not apply. Therefore, we employ a nonstandard but powerful Bracketing by Gaussian Hypotheses Theorem; 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 \(E[(\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u))^2] = 0.\) For this reason, our estimator \(\hat{\beta}(u)\) of \(\beta(u)\) is not consistent if \(N_g\) is bounded from above uniformly over \(g = 1, \ldots, G\) 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 $\hat{C}(\cdot, \cdot)$ that is defined for all $u_1, u_2 \in \mathcal{U}$ as

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

where

$$\hat{J}(u_1, u_2) = \frac{1}{G} \sum_{g=1}^{G} \left( (\hat{\alpha}_{g,1}(u_1) - x'_g\hat{\beta}(u_1))(\hat{\alpha}_{g,1}(u_2) - x'_g\hat{\beta}(u_2)) w_tw'_g \right),$$

$$\hat{S} = (\hat{Q}_{ww}\hat{Q}_{w}^{-1}\hat{Q}_{xw}^{-1}\hat{Q}_{ww})^{-1}\hat{Q}_{ww}\hat{Q}_{wxw},$$

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

**THEOREM 2**—Estimating $C$: Let Assumptions 1–8 hold. Then $\|\hat{C}(u_1, u_2) - C(u_1, u_2)\| = o_p(1)$ 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

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

Importantly for applied researchers, Theorems 1 and 2 demonstrate that heteroscedasticity-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. Traditional approaches to clustered standard errors are also valid in this setting; extending Theorems 1 and 2 to apply to settings with clustering is straightforward, but requires additional notation, and therefore we present these results in Appendix E of the Supplemental Material. As highlighted above, clustering in this context refers to clustering across groups. For example, if a group is state-by-year cell, the researcher could cluster at the state level.
6. THE EFFECT OF CHINESE IMPORT COMPETITION ON THE LOCAL WAGE DISTRIBUTION

6.1. Background on Wage Inequality

Over the past 40 years, wage inequality within the United States has increased drastically. 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. Recent work in Autor, Dorn, and Hanson (2013) focused on import competition and its effects on wages and employment in U.S. local labor markets. ADH studied the period 1990–2007, when the share of U.S. 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 U.S. labor markets that 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.

6.2. 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. The authors then estimated the following regression:

$\Delta \ln w_g = \beta_1 \Delta IPW_{g} + X'_g \beta_2 + \epsilon_g,$

where $\Delta \ln w_g$ 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

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

18 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).

19 The United States is covered exhaustively by 722 Commuting Zones (Tolbert and Sizer (1996)), each roughly corresponding to a local labor market.
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 \( \Delta IPW_g^U \), which represents the decadal change in Chinese imports per U.S. worker for the CZ and decade corresponding to group \( g \).^{20}

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 \( \Delta IPW_g^O \), a measure of import exposure that replaces the change in Chinese imports to the United States 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 to 2000 and 1.4% from 2000 to 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).^{21}

6.3. 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 \( \Delta \log w_g \), the change in the average log weekly wage in equation (9), with \( \Delta u \)-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.\(^{22}\) We instrument for \( \Delta IPW_g^U \) using \( \Delta IPW_g^O \) as described above. Recall that existing methods for handling endogeneity in quantile models are suited for the case where the

---

20 ADH apportion national industry-level import changes to local imports per worker using the weighted average of industry-level changes in the value of Chinese imports to the United States, with weights corresponding to the beginning-of-decade employment share of each industry in each CZ.

21 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 do not observe wages for the unemployed population.

22 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 of the Supplemental Material, 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.
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 \((\beta_1 \text{ in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group } g, \Delta \ln w_g, \text{ is replaced with the change in the } u\text{-quantile of log wages } \Delta \ln w_u^g). \text{ The dashed horizontal line is the ADH estimate of } \beta_1 \text{ in equation (9). 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.}

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.
FIGURE 2.—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 $1,000 increase in Chinese imports per worker on the female conditional wage distribution ($\beta_1$ in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group $g$, $\Delta \ln w_{g}$, is replaced with the change in the $u$-quantile of log wages $\Delta \ln w_{u}^g$). The dashed horizontal line is the ADH estimate of $\beta_1$ in equation (9). 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.

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. 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.

$^{23}$A coefficient of $-1.4$ log points, for example for the lower quantiles of Figure 1, corresponds to a 2.6% decrease in wages from 2000 to 2007 for the average commuting zone’s change in Chinese import exposure.
FIGURE 3.—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 $1,000 increase in Chinese imports per worker on the male conditional wage distribution ($\beta_1$ in equation (9) in the text when the change in average log wages for the commuting zone and decade corresponding to group $g$, $\Delta \ln w^g$, is replaced with the change in the $u$-quantile of log wages $\Delta \ln w^u_g$). The dashed horizontal line is the ADH estimate of $\beta_1$ in equation (9). 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.

7. CONCLUSION

In this paper, we present a quantile extension of Hausman and Taylor (1981), modeling the distributional effects of an endogenous group-level treatment. We develop an estimator, which we refer to as grouped IV quantile regression, and show that the estimator, as well as its standard errors, are easy to compute. We demonstrate that, in contrast to standard quantile regression, this estimator is asymptotically unbiased in the presence of the group-level shocks that are ubiquitous in applied microeconomic models. We illustrate the model and estimator with examples from labor, education, industrial organization, and urban economics. An empirical application to the setting of Autor, Dorn, and Hanson (2013) highlights the usefulness of our approach by estimating the effects of Chinese import competition on the distribution of wages—insights which would be missed by focusing on average effects alone. We believe the estimator has the potential for widespread practical use in applied microeconomics.
REFERENCES


Dept. of Economics, UCLA, 315 Portola Plaza, Los Angeles, CA 90095, U.S.A.; chetverikov@econ.ucla.edu,

Dept. of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305, U.S.A. and NBER; bjlarsen@stanford.edu,

and

Haas School of Business, University of California at Berkeley, Berkeley, CA 94720, U.S.A.; cjpalmer@berkeley.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received December, 2013; final revision received August, 2015.
APPENDIX A: SIMULATIONS

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

\[ y_{ig} = z_{ig} \gamma(u_{ig}) + \delta(u) + x_{g} \beta(u_{ig}) + \epsilon_{g}(u_{ig}), \]

\[ x_{g} = \pi w_{g} + \eta_{g} + \nu_{g}, \]

\[ \epsilon_{g}(u) = u \eta_{g} - u^2, \]

where \( w_{g}, \nu_{g}, \) and \( z_{ig} \) are each distributed \( \text{exp}(0.25 \cdot N[0, 1]) \); \( u_{ig} \) and \( \eta_{g} \) are both distributed \( U[0, 1] \); and random variables \( w_{g}, \nu_{g}, z_{ig}, u_{ig}, \) and \( \eta_{g} \) are mutually independent. Note that the form \( \epsilon_{g}(u) = u \eta_{g} - u^2 \) implies \( E[\epsilon_{g}(u)|w_{g}] = E[u \eta_{g} - u^2|w_{g}] = E[u^2 - u^2|w_{g}] = u - u = 0 \). 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 (10)–(12). The first case is exactly as in (10)–(12), with the group-level treatment of interest, \( x_{g} \), being endogenous (correlated with \( \epsilon_{g} \) through \( \eta_{g} \)). 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 \( \epsilon_{g} \)). In the second case, \( x_{g} \) is exogenous, where we set \( x_{g} = w_{g} \) in (11). 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 \( \epsilon_{g} \). In the third case, \( x_{g} \) is exogenous and no group-level unobservables are included, where we set \( x_{g} = w_{g} \) and \( \epsilon_{g} = 0 \). 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 \((N, G) = (25, 25), (200, 25), (25, 200), (200, 200)\). One thousand Monte Carlo replications were used. The results are displayed in Table A.I. Each panel displays the bias from the procedure for each decile \((u = 0.1, \ldots, 0.9)\) as well as the average absolute value of that bias, averaged over the nine deciles.

The top panel of Table A.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 \( N \)
**TABLE A.1**

**BIAS OF GROUPED IV QUANTILE REGRESSION VERSUS STANDARD QUANTILE REGRESSION**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(N, G) = (25, 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.316</td>
<td>0.042</td>
<td>−0.055</td>
<td>0.040</td>
<td>−0.007</td>
<td>0.038</td>
</tr>
<tr>
<td>0.2</td>
<td>0.447</td>
<td>0.076</td>
<td>0.015</td>
<td>0.078</td>
<td>−0.003</td>
<td>0.077</td>
</tr>
<tr>
<td>0.3</td>
<td>0.548</td>
<td>0.116</td>
<td>−0.024</td>
<td>0.116</td>
<td>−0.044</td>
<td>0.117</td>
</tr>
<tr>
<td>0.4</td>
<td>0.632</td>
<td>0.155</td>
<td>−0.128</td>
<td>0.154</td>
<td>−0.031</td>
<td>0.154</td>
</tr>
<tr>
<td>0.5</td>
<td>0.707</td>
<td>0.194</td>
<td>−0.182</td>
<td>0.193</td>
<td>−0.023</td>
<td>0.192</td>
</tr>
<tr>
<td>0.6</td>
<td>0.775</td>
<td>0.236</td>
<td>−0.192</td>
<td>0.233</td>
<td>−0.039</td>
<td>0.228</td>
</tr>
<tr>
<td>0.7</td>
<td>0.837</td>
<td>0.273</td>
<td>−0.161</td>
<td>0.270</td>
<td>−0.067</td>
<td>0.267</td>
</tr>
<tr>
<td>0.8</td>
<td>0.894</td>
<td>0.312</td>
<td>−0.106</td>
<td>0.311</td>
<td>−0.056</td>
<td>0.306</td>
</tr>
<tr>
<td>0.9</td>
<td>0.949</td>
<td>0.365</td>
<td>−0.106</td>
<td>0.361</td>
<td>−0.060</td>
<td>0.360</td>
</tr>
<tr>
<td>Avg. abs. bias</td>
<td><strong>0.197</strong></td>
<td><strong>0.108</strong></td>
<td><strong>0.195</strong></td>
<td><strong>0.037</strong></td>
<td><strong>0.193</strong></td>
<td><strong>0.008</strong></td>
</tr>
<tr>
<td>(N, G) = (200, 25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.316</td>
<td>0.005</td>
<td>0.010</td>
<td>−0.004</td>
<td>−0.016</td>
<td>0.002</td>
</tr>
<tr>
<td>0.2</td>
<td>0.447</td>
<td>0.005</td>
<td>0.007</td>
<td>0.001</td>
<td>−0.010</td>
<td>0.002</td>
</tr>
<tr>
<td>0.3</td>
<td>0.548</td>
<td>0.006</td>
<td>−0.006</td>
<td>0.006</td>
<td>−0.012</td>
<td>0.003</td>
</tr>
<tr>
<td>0.4</td>
<td>0.632</td>
<td>0.011</td>
<td>−0.021</td>
<td>0.007</td>
<td>−0.010</td>
<td>0.005</td>
</tr>
<tr>
<td>0.5</td>
<td>0.707</td>
<td>0.008</td>
<td>−0.039</td>
<td>0.008</td>
<td>−0.002</td>
<td>0.007</td>
</tr>
<tr>
<td>0.6</td>
<td>0.775</td>
<td>0.004</td>
<td>−0.021</td>
<td>0.009</td>
<td>−0.004</td>
<td>0.009</td>
</tr>
<tr>
<td>0.7</td>
<td>0.837</td>
<td>0.006</td>
<td>−0.011</td>
<td>0.007</td>
<td>−0.003</td>
<td>0.009</td>
</tr>
<tr>
<td>0.8</td>
<td>0.894</td>
<td>−0.010</td>
<td>−0.007</td>
<td>−0.011</td>
<td>−0.001</td>
<td>−0.011</td>
</tr>
<tr>
<td>0.9</td>
<td>0.949</td>
<td>−0.031</td>
<td>0.008</td>
<td>−0.038</td>
<td>0.003</td>
<td>−0.028</td>
</tr>
<tr>
<td>Avg. abs. bias</td>
<td><strong>0.010</strong></td>
<td><strong>0.017</strong></td>
<td><strong>0.010</strong></td>
<td><strong>0.007</strong></td>
<td><strong>0.009</strong></td>
<td><strong>0.014</strong></td>
</tr>
<tr>
<td>(N, G) = (25, 200)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.316</td>
<td>0.002</td>
<td>0.019</td>
<td>0.001</td>
<td>−0.006</td>
<td>0.000</td>
</tr>
<tr>
<td>0.2</td>
<td>0.447</td>
<td>0.008</td>
<td>0.009</td>
<td>0.003</td>
<td>−0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>0.3</td>
<td>0.548</td>
<td>0.005</td>
<td>−0.023</td>
<td>0.004</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>0.4</td>
<td>0.632</td>
<td>0.007</td>
<td>−0.015</td>
<td>0.004</td>
<td>−0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>0.5</td>
<td>0.707</td>
<td>0.005</td>
<td>−0.027</td>
<td>0.000</td>
<td>−0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>0.6</td>
<td>0.775</td>
<td>0.004</td>
<td>−0.037</td>
<td>0.001</td>
<td>−0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>0.7</td>
<td>0.837</td>
<td>0.003</td>
<td>−0.027</td>
<td>0.000</td>
<td>−0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>0.8</td>
<td>0.894</td>
<td>0.000</td>
<td>−0.022</td>
<td>0.000</td>
<td>−0.003</td>
<td>0.001</td>
</tr>
<tr>
<td>0.9</td>
<td>0.949</td>
<td>−0.003</td>
<td>−0.023</td>
<td>0.000</td>
<td>−0.003</td>
<td>−0.001</td>
</tr>
<tr>
<td>Avg. abs. bias</td>
<td><strong>0.004</strong></td>
<td><strong>0.023</strong></td>
<td><strong>0.002</strong></td>
<td><strong>0.004</strong></td>
<td><strong>0.001</strong></td>
<td><strong>0.004</strong></td>
</tr>
</tbody>
</table>

*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 $u \in [0.1, \ldots , 0.9]$ as well as the average absolute value of the bias, averaged over the nine deciles.*
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—$\alpha_g$ 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 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 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 $\alpha_g$ 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).

\section*{APPENDIX B: 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.

\(^{24}\)With $G > 200$, the computation time ratio drastically increases further, with standard optimization packages often failing to converge appropriately.
THEOREM 3—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 $g = 1, \ldots, G$ and $x \in (0, \bar{c})$,

\[
P\left( \sup_{u \in U} \left\| \hat{\alpha}_g(u) - \alpha_g(u) \right\| > x \right) \leq Ce^{-cx^2Ng}.
\]

REMARK 3: The bound provided in Theorem 3 is non-asymptotic. In principle, it is also possible to calculate the exact constants in the inequality (13). We do not calculate these constants because they are not needed for our results. Since $\hat{\alpha}_g(u)$ is the classical Koenker and Bassett’s (1978) quantile regression estimator of $\alpha_g(u)$, Theorem 3 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.

APPENDIX C: UNIFORM CONFIDENCE INTERVALS

In this section, we show how to obtain confidence bands for $\beta(u)$ that hold uniformly over $U$. Observe that $\beta(u)$ is a $d_x$-vector, that is, $\beta(u) = (\beta_1(u), \ldots, \beta_{d_x}(u))^\prime$. Without loss of generality, we focus on $\beta_1(u)$, the first component of $\hat{\beta}(u)$. Let $\hat{\beta}_1(u), V(u)$, and $\hat{V}(u)$ denote the first component of $\hat{\beta}(u)$, the $(1, 1)$ component of $C(u, u)$, and the $(1, 1)$ component of $\hat{C}(u, u)$, respectively. Define

\[
T = \sup_{u \in U} \sqrt{G} \left| \hat{V}(u)^{-1/2} (\hat{\beta}_1(u) - \beta_1(u)) \right|,
\]

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

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

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 $\epsilon_1, \ldots, \epsilon_G$ be an i.i.d. sequence of $N(0, 1)$ random variables that are independent of the data. Also, let $\hat{w}_{g, 1}$ denote the first component of the vector $\hat{S}w_g$. Then the multiplier bootstrap statistic is

\[
T_{MB} = \sup_{u \in U} \frac{1}{\sqrt{G\hat{V}(u)}} \sum_{g=1}^G (\epsilon_g (\hat{\alpha}_{g, 1}(u) - x'_g \hat{\beta}(u)) \hat{w}_{g, 1}).
\]

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 (15) 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 4**—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

\[
P\left( \beta_1(u) \in \left[ \hat{\beta}_1(u) - \hat{c}_{1-\alpha} \sqrt{\frac{\hat{V}(u)}{G}}, \hat{\beta}_1(u) + \hat{c}_{1-\alpha} \sqrt{\frac{\hat{V}(u)}{G}} \right] \right)
\]

for all \( u \in \mathcal{U} \) \( \rightarrow 1 - \alpha \)
as \( G \rightarrow \infty \).

**Remark 4:** Uniform confidence bands are typically larger than the point-wise confidence bands based on the result (8). 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.

**Appendix D: Joint Inference on Group-Specific Effects**

In this section, we are concerned with inference on group-specific effects \( \alpha_{g,1}(u), g = 1, \ldots, G \), in the model (2)–(3) defined in Section 2. In particular, we are interested in constructing the confidence bands \( [\hat{\alpha}_{g,1}^l, \hat{\alpha}_{g,1}^r] \) for \( \alpha_{g,1}(u) \) that are adjusted for multiplicity of the effects, that is, we would like to have the bands satisfying

\[
P(\alpha_{g,1}(u) \in [\hat{\alpha}_{g,1}^l, \hat{\alpha}_{g,1}^r] \text{ for all } g = 1, \ldots, G) \rightarrow 1 - \alpha.
\]

Thus, the confidence bands \( [\hat{\alpha}_{g,1}^l, \hat{\alpha}_{g,1}^r] \) cover the true group-specific effects \( \alpha_{g,1} \) for all \( g = 1, \ldots, G \) simultaneously with probability approximately \( 1 - \alpha \).

The main challenge here is that we have \( G \) parameters \( \alpha_{g,1}(u), g = 1, \ldots, G \), and only \( N_g \) observations to estimate \( \alpha_{g,1} \), where \( N_g \) is potentially smaller than \( G \) (recall that we impose Assumption 3, according to which \( G^{2/3}(\log N_G)/N_G \rightarrow 0 \) as \( G \rightarrow \infty \) where \( N_G = \min_{g=1,\ldots,G} N_g \)). 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 $u(1 - u) J_g(u)^{-1} E_g[z_{g,1} z_{g,1}'] J_g(u)^{-1}$; 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

$$(17) \quad P(\alpha_{g,1}(u) \in \left[ \hat{\alpha}_{g,1}(u) - c_{1-\alpha} \sqrt{I_g/N_g}, \hat{\alpha}_{g,1}(u) + c_{1-\alpha} \sqrt{I_g/N_g} \right]) \to 1 - \alpha \quad \text{as} \quad N_g \to \infty.$$ 

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 (17) 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 (17), however, is that they do not take into account multiplicity of the effects $\alpha_{g,1}(u)$, $g = 1, \ldots, G$. This is especially important given that the assumptions, max$_{1 \leq g \leq G}$ $|Y_g|$ where $Y_1, \ldots, Y_G$ 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 5—Joint Inference on Group-Specific Effects:** Let Assumptions 1–8 hold. In addition, suppose that $I_g \geq c_M$ for all $g = 1, \ldots, G$ and $\bar{N}_G/N_G \leq C_M$ where $N_G = \min_{1 \leq g \leq G} N_g$ and $\bar{N}_G = \max_{1 \leq g \leq G} N_g$. Let $c_{1-\alpha}^M$ be the $(1 - \alpha)$ quantile of max$_{1 \leq g \leq G} |Y_g|$ where $Y_1, \ldots, Y_G$ are i.i.d. $N(0, 1)$ random variables. Then

$$P(\alpha_{g,1}(u) \in \left[ \hat{\alpha}_{g,1}(u) - c_{1-\alpha}^M \sqrt{I_g/N_g}, \hat{\alpha}_{g,1}(u) + c_{1-\alpha}^M \sqrt{I_g/N_g} \right]) \to 1 - \alpha$$

for all $g = 1, \ldots, G$ as $G \to \infty$.

**REMARK 5:** We note that the size of the bands in this theorem, max$_{1 \leq g \leq G} 2 \times c_{1-\alpha}^M (I_g/N_g)^{1/2}$, is shrinking to zero as $G$ gets large. Indeed, under our assumptions, max$_{1 \leq g \leq G} I_g \leq C$ for some constant $C$, which is independent of $G$. In addition, $c_{1-\alpha}^M \leq (C \log G)^{1/2}$ for some absolute constant $C$. Therefore, max$_{1 \leq g \leq G} c_{1-\alpha}^M (I_g/N_g)^{1/2} \leq (C \log G/N_G)^{1/2} \to 0$ by our growth condition in Assumption 3 (for some possibly different constant $C$).
APPENDIX E: CLUSTERED STANDARD ERRORS

In this section, we consider the model from the main text, which is defined in equations (2)–(3), 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, 2, and 4.

Before presenting these results, we first provide several examples of where this clustering would be useful; referencing the examples in Section 4, a group is a grade-by-school-by-year cell, and the researcher may be interested in clustering at the school or school-by-grade level, for example. In Example 2, a group is a state-by-year combination, and the researcher may be interested in clustering at the state level. In Example 3, a group is a given MSA, and the researcher may be interested in clustering at the region level (where a region contains several MSAs). In Example 4, a group is a market-by-time-period combination, and the researcher may be interested in clustering at the market level.

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

ASSUMPTION 1′—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 $(x_g, \alpha_g)$. (iii) For each $m = 1, \ldots, M$, the number of elements in the set $C_G(m)$ 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:

ASSUMPTION 6′—Noise: (i) For all $g = 1, \ldots, G$, $E[\sup_{u \in U} |\epsilon_g(u)|^{4+CM}] \leq C_M$. (ii) For some (matrix-valued) function $J^{CS}: U \times U \to \mathbb{R}^{d_u \times d_u}$,

$$
\frac{1}{G} \sum_{m=1}^{M} E\left[ \left( \sum_{g \in C_G(m)} \epsilon_g(u_1)w_g \right) \left( \sum_{g \in C_G(m)} \epsilon_g(u_2)w'_g \right) \right] \to J^{CS}(u_1, u_2)
$$

uniformly over $u_1, u_2 \in U$. (iii) For all $u_1, u_2 \in U$, $|\epsilon_g(u_2) - \epsilon_g(u_1)| \leq C_L|u_2 - u_1|$. 


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 $C_G(m)$ 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 $\beta(u), u \in U$, 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

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

where $G_{CS}(\cdot)$ is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function $C_{CS}(u_1, u_2) = SJ_{CS}(u_1, u_2)S'$, where $S = (Q_{xw}Q_{xw}'Q_{xw}W) - Q_{xw}'Q_{xw}$, $Q_{xw}$ 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 $C_{CS}(\cdot, \cdot)$ of the limiting Gaussian process $G_{CS}(\cdot)$. We suggest estimating $C_{CS}(\cdot, \cdot)$ by $\hat{C}_{CS}(\cdot, \cdot)$ defined for all $u_1, u_2 \in U$ as

$$\hat{C}_{CS}(u_1, u_2) = \hat{S}\hat{j}_{CS}(u_1, u_2)\hat{S},$$

where

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

$$\times \left( \sum_{g \in C_G(m)} (\hat{\alpha}_{g,2}(u_2) - x_g'\hat{\beta}(u_2))w_g' \right),$$

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

**Theorem 7—Estimating $C_{CS}$ Under Cluster Sampling:** Let Assumptions 1’, 2–5, 6’, 7, and 8 hold. Then $\|\hat{C}_{CS}(u_1, u_2) - C_{CS}(u_1, u_2)\| = o_p(1)$ uniformly over $u_1, u_2 \in U$.

Finally, we show how to obtain confidence bands for $\beta(u)$ that hold uniformly over $U$. Observe that $\beta(u)$ is a $d_x$-vector, that is, $\beta(u) = (\beta_1(u), \ldots, \beta_{d_x}(u))$. 

As before, we focus on $\beta_1(u)$, the first component of $\beta(u)$, and we suggest constructing uniform confidence bands via a multiplier bootstrap method. An important difference from the results with no clustering, however, 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 $C^{CS}(u, u)$, and the $(1, 1)$st component of $\hat{C}^{CS}(u, u)$, respectively. Define

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

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 $\epsilon_1, \ldots, \epsilon_M$ be an i.i.d. sequence of $N(0, 1)$ random variables that are independent of the data. Also, let $\hat{w}_{g,1}$ denote the first component of the vector $\hat{w}_g$. Then the multiplier bootstrap statistic is

$$T_{MB} = \sup_{u \in U} \frac{1}{\sqrt{G \hat{V}(u)}} \sum_{m=1}^{M} \epsilon_m \left( \sum_{g \in \hat{C}_G(m)} (\hat{\alpha}_{g,1}(u) - x'_g \hat{\beta}(u)) \hat{w}_{g,1} \right).$$

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 $u \in U$. Then

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

for all $u \in U$ \(\rightarrow\) $1 - \alpha$

as $G \rightarrow \infty$.

**Appendix F: Proofs**

In this appendix, we first prove some preliminary lemmas. Then we present the proofs of Theorems 1–5 stated in the main text and in Appendices B–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

\[
A(u) = \left( \alpha_{1,1}(u), \ldots, \alpha_{G,1}(u) \right)',
\]
\[
\tilde{\beta}(u) = \left( X'P_W X \right)^{-1}(X'P_W A(u)),
\]
\[
J_g(u) = E_g\left[ z_{1g} z_{1g} f_g(z_{1g} \alpha_g(u)) \right].
\]

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

\[
f_{\eta,\alpha,u}(z, y) = (z' \eta) \cdot (1\{y \leq z' \alpha\} - u).
\]

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 \( \eta, \alpha \) 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 \( h_{\alpha,u} : \mathbb{R}^{d_z} \times \mathbb{R} \rightarrow \mathbb{R}^{d_z} \) be defined by

\[
h_{\alpha,u}(z, y) = z(1\{y \leq z' \alpha\} - u),
\]

and let \( h_{k,\alpha,u} \) denote \( k \)th 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 \( k = 1, \ldots, d_z \).

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

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

for \( f \in \mathcal{F}, \mathcal{H}, \) or \( \mathcal{H}_k, k = 1, \ldots, d_z \).

**Preliminary Lemmas**

In all lemmas, we implicitly impose Assumptions 1–8.

**Lemma 1:** As \( G \rightarrow \infty \),

\[
\hat{Q}_{xw} = \frac{1}{G} \sum_{g=1}^{G} x_g w_g' \rightarrow_p Q_{xw},
\]

\[
\hat{Q}_{ww} = \frac{1}{G} \sum_{g=1}^{G} w_g w_g' \rightarrow_p Q_{ww},
\]

where \( Q_{xw} \) and \( Q_{ww} \) appear in Assumption 2.
PROOF: We only prove (21). The proof of (22) is similar. To prove (21), 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

\[
\frac{1}{G} \sum_{g=1}^{G} (x_g w'_g - E[x_g w'_g]) \to_p 0. \tag{23}
\]

In turn, (23) follows from Assumptions 2(iv) and 4(i) and Chebyshev’s inequality. Hence, (21) follows. This completes the proof of the lemma. \( Q.E.D. \)

**Lemma 2:** As \( G \to \infty \),

\[
\frac{1}{G} \sum_{g=1}^{G} \epsilon_g(u_1) \epsilon_g(u_2) w_g w'_g \to_p J(u_1, u_2)
\]

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

\[
\frac{1}{G} \sum_{g=1}^{G} E[\epsilon_g(u_1) \epsilon_g(u_2) w_g w'_g] \to J(u_1, u_2)
\]

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

\[
\frac{1}{G} \sum_{g=1}^{G} (\epsilon_g(u_1) \epsilon_g(u_2) w_{g,k} w_{g,l} - E[\epsilon_g(u_1) \epsilon_g(u_2) w_{g,k} w_{g,l}]) \to_p 0 \tag{24}
\]

uniformly over \( u_1, u_2 \in \mathcal{U} \) for all \( k, l = 1, \ldots, d_w \).

To this end, fix \( u_1, u_2 \in \mathcal{U} \) and \( k, l = 1, \ldots, d_w \). We first show (24) for these values of \( u_1, u_2, k, \) and \( l \). Note that we cannot use Chebyshev’s inequality here because \( E[(\epsilon_g(u_1) \epsilon_g(u_2) w_{g,k} w_{g,l})^2] \) 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,

\[
E[|\epsilon_g(u_1) \epsilon_g(u_2) w_{g,k} w_{g,l}|^{1+\delta}] \leq (E[|\epsilon_g(u_1) \epsilon_g(u_2)|^{2+2\delta}] \cdot E[|w_{g,k} w_{g,l}|^{2+2\delta}])^{1/2}.
\]
In turn,
\[ E[|\epsilon_g(u_1)\epsilon_g(u_2)|^{2+\delta}] \leq E\left[ \sup_{u \in D}|\epsilon_g(u)|^{4+4\delta} \right] \leq C_M, \]
\[ E[|w_{g,k}w_{g,l}|^{2+\delta}] \leq E\left[ \|w_g\|^{4+\delta} \right] \leq C_M, \]
by Assumptions 6(i) and 2(iv). Hence,
\[ E\left[ |\epsilon_g(u_1)\epsilon_g(u_2)w_{g,k}w_{g,l}|^{1+\delta} \right] \leq C_M, \]
and so denoting \( X_g = \epsilon_g(u_1)\epsilon_g(u_2)w_{g,k}w_{g,l} - E[\epsilon_g(u_1)\epsilon_g(u_2)w_{g,k}w_{g,l}] \), we obtain
\[ (25) \quad E[|X_g|^{1+\delta}] \leq C. \]

With this notation, (24) 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,

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

where in the last inequality we used (25). Hence, by Markov’s inequality, for any \( \epsilon > 0 \),

\[
P\left( \left| \frac{1}{G} \sum_{g=1}^{G} X_{g,>N} \right| > \epsilon \right) \leq \frac{1}{\epsilon G} \sum_{g=1}^{G} E[X_{g,>N}] \leq \frac{C}{\epsilon N^\delta},
\]
and since \( |E[X_{g,\leq N}]| = |E[X_{g,>N}]| \leq CN^{-\delta} \),

\[
P\left( \left| \frac{1}{G} \sum_{g=1}^{G} X_{g,\leq N} \right| > \epsilon + CN^{-\delta} \right)
\]
\[
\leq P\left( \left| \frac{1}{G} \sum_{g=1}^{G} (X_{g,\leq N} - E[X_{g,\leq N}]) \right| > \epsilon \right) \]
\[
\leq \frac{1}{\varepsilon^2 G^2} \sum_{g=1}^{G} E[X_g^2 \leq N]
\leq \frac{N^2}{\varepsilon^2 G}.
\]

Thus, setting \( N = G^{1/3} \), we obtain \( G^{-1} \sum_{g=1}^{G} X_g \to p 0 \), which is equivalent to (24) for given \( u_1, u_2, k, \) and \( l \).

Next, to show that (24) holds uniformly over \( u_1, u_2 \in U \), for \( \delta > 0 \), let \( U_\delta \) be a finite subset of \( U \) such that, for any \( u \in U \), there exists \( u' \in U_\delta \) satisfying \( |\varepsilon_g(u) - \varepsilon_g(u')| \leq \delta \). Existence of such a set \( U_\delta \) follows from Assumption 6(iii).

Then
\[
\sup_{u_1, u_2 \in U} \left| \frac{1}{G} \sum_{g=1}^{G} (\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l} - E[\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}]) \right|
\leq \max_{u_1, u_2 \in U_\delta} \left| \frac{1}{G} \sum_{g=1}^{G} (\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l} - E[\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}]) \right|
\]
\[
+ \frac{2\delta}{G} \sum_{g=1}^{G} \left( \sup_{u \in \mathcal{U}} |\varepsilon_g(u)| \cdot |w_{g,k} w_{g,l}| + E[\sup_{u \in \mathcal{U}} |\varepsilon_g(u)| \cdot |w_{g,k} w_{g,l}|] \right)
\]
\[
= o_p(1) + \delta \cdot O_p(1)
\]
by the result above and Chebyshev’s inequality. Since \( \delta \) is arbitrary, this completes the proof. \( \square \)

**LEMMA 3:** As \( G \to \infty \),
\[
\frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \varepsilon_g(\cdot) \Rightarrow \xi_0^0(\cdot), \quad \text{in} \quad \ell^\infty(\mathcal{U}),
\]
where \( \xi_0^0 \) 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, e.g., Theorems 11.2.5 and 11.2.3 in Lehmann and Romano (2005)) that
\[
\left( \frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \varepsilon_g(u) \right)_{u \in \mathcal{U}'} \Rightarrow \left( N(u) \right)_{u \in \mathcal{U}'},
\]
where \( \left( N(u) \right)_{u \in \mathcal{U}'} \) is a zero-mean Gaussian vector with covariance function \( J(u_1, u_2) \) for all \( u_1, u_2 \in \mathcal{U}' \). Therefore, to prove the asserted claim, we can
apply Theorem 14. In particular, it suffices to verify conditions (i)–(iii) of Theorem 14 with $Z_g(u) = G^{-1/2}w_{g,k}e_g(u)$, $g = 1, \ldots, G$ and $u \in \mathcal{U}$, for all $k = 1, \ldots, d_w$. In the verification, we will use the Gaussian-dominated semi-metric $\rho : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ defined by $p(u_1, u_2) = C|u_2 - u_1|$ 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 $\delta = 1 + c_M/2$,

$$\sum_{g=1}^{G} E\left[\sup_{u \in \mathcal{U}} |Z_g(u)| \cdot 1\{\sup_{u \in \mathcal{U}} |Z_g(u)| > \eta\}\right]$$

$$\leq \frac{1}{\eta^\delta G^{1/2+\delta/2}} \sum_{g=1}^{G} E\left[\sup_{u \in \mathcal{U}} |\epsilon_g(u)|^{1+\delta}|w_{g,k}|^{1+\delta}\right]$$

$$\leq \frac{1}{\eta^\delta G^{1/2+\delta/2}} \sum_{g=1}^{G} \left( E\left[\sup_{u \in \mathcal{U}} |\epsilon_g(u)|^{2+2\delta}\right] \cdot E\left[|w_{j,k}|^{2+2\delta}\right]\right)^{1/2} \to 0$$

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

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

$$\sum_{g=1}^{G} E\left[(Z(u_2) - Z(u_1))^2\right] = \frac{1}{G} \sum_{g=1}^{G} E\left[(w_{g,k}e_g(u_2) - w_{g,k}e_g(u_1))^2\right]$$

$$\leq \frac{C}{G} \sum_{g=1}^{G} E\left[w_{g,k}^2|u_2 - u_1|^2\right]$$

$$\leq C|u_2 - u_1|^2 \leq \rho^2(u_1, u_2)$$

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 $\epsilon > 0$,

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

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

$$\leq C \sup_{p(u_1,u_2) \leq 2\epsilon} |u_2 - u_1|^2 \leq \epsilon^2$$
by Assumptions 2(iv) and 6(iii) since the constant $C$ in the definition of $ho(u_1, u_2)$ is large enough. The asserted claim follows from an application of Theorem 14. Q.E.D.

**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 $\|\alpha\| = 1$,

$$\alpha'J_g(u)\alpha \geq c_f\alpha' E_g[z_{1g}z'_{1g}]\alpha \geq c_f c_M,$$  

(26) 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 (26). 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

$$\|J_g(u_2) - J_g(u_1)\| \leq \|E_g[z_{1g}z'_{1g} \cdot f_g(z'_{1g}\alpha_g(u_2)) - f_g(z'_{1g}\alpha_g(u_1))]| |$$

$$\leq C_f C_M C_L |u_2 - u_1| \cdot \|E_g[z_{1g}z'_{1g}]\|$$

$$\leq C_f C_M^3 C_L |u_2 - u_1|,$$

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

$$\|J_g(u_2) - J_g(u_1)\| \leq \|J_g(u_1)\| + \|J_g(u_2)\| \leq 2C_f \|E_g[z_{1g}z'_{1g}]\|$$

$$\leq 2C_f C_M^2 \leq c_f^{-1} C_f C_M^3 C_L |u_2 - u_1|,$$

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. Q.E.D.
Lemma 5: There exist constants \( c, C > 0 \) such that, for all \( g = 1, \ldots, G \),

\[
\begin{aligned}
\| E_g[h_{a,u}(z_{1g}, y_{1g})] - J_g(u)(\alpha - \alpha_g(u)) \| &\leq C \| \alpha - \alpha_g(u) \|^2, \\
E_g\left[ (\alpha - \alpha_g(u))' h_{a,u}(z_{1g}, y_{1g}) \right] &\geq c \| \alpha - \alpha_g(u) \|^2,
\end{aligned}
\]

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

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

where \( r_n(u) \) is the remainder and \( F_g(\cdot) \) is the conditional distribution function of \( y_{1g} \) given \((z_{1g}, x_{g}, \alpha_g)\). 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 \( u \)th 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. Q.E.D.

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

\[
\{ (z, y, t) : f_{\eta,a,u}(z, y) > t \} = \left( \{ y \leq z' \alpha \} \cap \{ z' \eta > t/(1 - u) \} \right) \\
\cup \left( \{ y > z' \alpha \} \cap \{ z' \eta < -t/u \} \right).
\]

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 \( \mathcal{H}_k \subset \mathcal{F} \). Q.E.D.
**Lemma 7:** For any $\varphi \geq 1$, there exists a constant $C > 0$ such that, for all $g = 1, \ldots, G$,

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

**Proof:** Observe that

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

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 $E_g \left[ \sup_{f \in \mathcal{H}_k} \left\| \mathcal{H}^g(f) \right\| \right] \leq C$, and so Theorem 13 shows that

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

The asserted claim follows. \( Q.E.D. \)

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

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

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

**Proof:** Fix some $u_1 \in U$. Observe that

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

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

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

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

$$
| h_{k, \alpha_g(u_2), u_2}(z_{ig}, y_{ig}) - h_{k, \alpha_g(u_1), u_1}(z_{ig}, y_{ig}) | \leq F(z_{ig}, y_{ig})
$$
by Assumption 4(i). Note that $E_g[F^2(\mathcal{I}_g, \mathcal{Y}_g)] \leq C \epsilon$ for some $C > 0$ by Assumption 7(ii) if $\epsilon \leq 1$. Also, for $M = \max_{1 \leq i \leq N_g} F(\mathcal{I}_g, \mathcal{Y}_g)$, we have $E[M^2] \leq C N_g \epsilon$.

Further, by Lemma 6, $\mathcal{H}_k$ is a VC subgraph class of functions, so that the function class $\mathcal{H}_k = \{h_{k, \alpha}(u) : u \in [u_1 - \epsilon, u_1 + \epsilon]\}$ is a VC type class by Theorem 9. So, applying Theorem 11 with $F$ as an envelope yields
\[
E_g\left[ \sup_{u_2 \in U : |u_2 - u_1| \leq \epsilon} \left| G_h(\alpha(u_2)) - G_h(\alpha(u_1)) \right|^2 \right] \leq C \epsilon,\
\]
and so Theorem 13 shows that
\[
E_g\left[ \sup_{u_2 \in U : |u_2 - u_1| \leq \epsilon} \left| G_h(\alpha(u_2)) - G_h(\alpha(u_1)) \right|^4 \right] \leq C \epsilon,\
\]
since
\[
E_g\left[ \max_{1 \leq i \leq N_g} \sup_{u_2 \in U : |u_2 - u_1| \leq \epsilon} \left| N_g^{-1/2}(h_{k, \alpha}(u_2), u_2) - h_{k, \alpha}(u_1, u_1) \right|^4 \right] \leq N_g^{-1} \max_{1 \leq i \leq N_g} E_g\left[ \sup_{u_2 \in U : |u_2 - u_1| \leq \epsilon} \left| (h_{k, \alpha}(u_2), u_2) - h_{k, \alpha}(u_1, u_1) \right|^4 \right] \leq N_g^{-1} E_g[F^4(\mathcal{I}_g, \mathcal{Y}_g)] \leq C \epsilon.\
\]
The asserted claim follows. \(Q.E.D.\)

**Lemma 9:** There exist constants $c, C > 0$ such that, for all $g = 1, \ldots, G$,
\[
E_g\left[ \sup_{u \in U} \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq \epsilon} \left| G_h(\alpha(u)) - G_h(\alpha_g(u)) \right|^2 \right] \leq C(\epsilon \log(1/\epsilon) + N_g^{-1} \log^2(1/\epsilon))
\]
for all $\epsilon \in (0, c)$.

**Proof:** Observe that
\[
E_g\left[ \sup_{u \in U} \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq \epsilon} \left| G_h(\alpha(u)) - G_h(\alpha_g(u)) \right|^2 \right] \leq C \sum_{k=1}^{d_z} E_g\left[ \sup_{u \in U} \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq \epsilon} \left| G_h(\alpha(u)) - G_h(\alpha_g(u)) \right|^2 \right].
\]
Consider the function class

\[ \tilde{H}_k = \{ h_{k,a,u} - h_{k,a_g(u),u} : u \in \mathcal{U}; \alpha \in \mathbb{R}^{d_z}; \| \alpha - \alpha_g(u) \| \leq \epsilon \} \]

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

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

and so Theorem 13 gives

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

The asserted claim follows. \( Q.E.D. \)

**Lemma 10:** Uniformly over \( u \in \mathcal{U} \),

\[
\frac{1}{\sqrt{G}} \sum_{g=1}^{G} J_g^{-1}(u) G^g(h_{a_g(u),u}) w_g' = O_p(1).
\]

**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, \( \rho \) is Gaussian-dominated; see discussion before Theorem 14 for the definition. Define \( v_g(u) = J_g^{-1}(u) G^g(h_{a_g(u),u}) \) 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 \( k \)th and \( m \)th components of \( v_g(u) \) and \( w_g \), respectively. Then the asserted claim is equivalent to the statement that

\[
\sum_{g=1}^{G} Z_{g,k,m}(u) = O_p(1) \quad \text{uniformly over} \quad u \in \mathcal{U}
\]
for all \( k \) and \( m \). To prove (31), 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 \( a > 0 \),

\[
(32) \quad \sum_{g=1}^{G} E \left[ \sup_{u \in \mathcal{U}} |Z_{g,k,m}(u)| \cdot 1 \{ \sup_{u \in \mathcal{U}} |Z_{g,k,m}(u)| > a \} \right]
\]

\[
\leq a^{-1} \sum_{g=1}^{G} E \left[ \sup_{u \in \mathcal{U}} Z_{g,k,m}^2(u) \cdot 1 \{ \sup_{u \in \mathcal{U}} |Z_{g,k,m}(u)| > a \} \right]
\]

\[
\leq \frac{1}{aG} \sum_{g=1}^{G} E \left[ \sup_{u \in \mathcal{U}} (v_{g,k}(u)w_{g,m})^2 \cdot 1 \{ \sup_{u \in \mathcal{U}} |v_{g,k}(u)w_{g,m}| > \sqrt{Ga} \} \right].
\]

Further, pick some \( 0 < \phi < 2 \). The expression under the sum in (32) is bounded from above by Lemma 4 by

\[
\frac{C}{a^\phi G^{\phi/2}} E \left[ \sup_{u \in \mathcal{U}} \| G^\phi(h_{\alpha g}(u), u) \|^{2+\phi} \| w_g \|^{2+\phi} \right]
\]

\[
\leq \frac{C}{a^\phi G^{\phi/2}} \left( E \left[ \sup_{u \in \mathcal{U}} \| G^\phi(h_{\alpha g}(u), u) \|^{4(2+\phi)/(2-\phi)} \right] \right)^{(2-\phi)/4} \left( E[\| w_g \|^4] \right)^{(2+\phi)/4}
\]

\[
\leq \frac{C}{a^\phi G^{\phi/2}} \to 0
\]

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 \( u_1, u_2 \in \mathcal{U} \),

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

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

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

\[
E_g \left[ (v_{g,k}(u_2) - v_{g,k}(u_1))^4 \right]
\]

\[
\leq CE_g \left[ \| J_{g}^{-1}(u_2) \|^4 \cdot \| G^\phi(h_{\alpha g}(u_2), u_2 - h_{\alpha g}(u_1), u_1) \|^4 \right]
\]

\[
+ CE_g \left[ \| J_{g}^{-1}(u_2) - J_{g}^{-1}(u_1) \|^4 \cdot \| G^\phi(h_{\alpha g}(u_1), u_1) \|^4 \right]
\]
\[\leq CE_g\left(\|G^a(h_{\alpha g(u_2)},u_2 - h_{\alpha g(u_1)},u_1)\|^4\right)\]
\[+ CE_g\left(\|G^a(h_{\alpha g(u_1)},u_1)\|^4\right) \cdot |u_2 - u_1|^4,\]

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

\[(33) \quad E_g\left[\|G^a(h_{\alpha g(u_2)},u_2 - h_{\alpha g(u_1)},u_1)\|^4\right] \leq C|u_2 - u_1| \quad \text{and} \quad E_g\left[\|G^a(h_{\alpha g(u_1)},u_1)\|^4\right] \leq C,
\]

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

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

and so

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

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, to verify condition (iii) of Theorem 14, observe that, for any \(\epsilon > 0\) and \(u_1 \in \mathcal{U}\),

\[\sup_{t>0} \sum_{g=1}^{G} t^2 P\left(\sup_{u_2 \in \mathcal{L}: p(u_1, u_2) \leq \epsilon} |Z_{g,k,m}(u_2) - Z_{g,k,m}(u_1)| > t\right) \leq \sum_{g=1}^{G} E\left[\sup_{u_2 \in \mathcal{L}: p(u_1, u_2) \leq \epsilon} |Z_{g,k,m}(u_2) - Z_{g,k,m}(u_1)|^2\right] \]
\[= \frac{1}{G} \sum_{g=1}^{G} E\left[\sup_{u_2 \in \mathcal{L}: p(u_1, u_2) \leq \epsilon} |v_{g,k}(u_2) - v_{g,k}(u_1)|^2 w_{g,m}^2\right] \leq \epsilon^2,
\]

where the second line follows from Markov’s inequality, and the last inequality follows by selecting sufficiently large constant \(C\) in the definition of \(\rho\) and using the same argument as that in verification of condition (ii) since the first inequality in (33) used in the verification of condition (ii) can be replaced by

\[E_g\left[\sup_{u_2 \in \mathcal{L}: p(u_1, u_2) \leq \epsilon} \|G^a(h_{\alpha g(u_2)},u_2 - h_{\alpha g(u_1)},u_1)\|^4\right] \leq c\epsilon^4
\]
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. Therefore, for any $\epsilon > 0$ and $u \in U$,

$$\sup_{t>0} \sum_{g=1}^{G} t^2 P \left( \sup_{u_1, u_2 \in U : \rho(u_1, u) \leq \epsilon, \rho(u_2, u) \leq \epsilon} \left| Z_{g,k,m}(u_2) - Z_{g,k,m}(u_1) \right| > t \right)$$

$$\leq 2 \sup_{t>0} \sum_{g=1}^{G} t^2 P \left( \sup_{u_1 \in U : \rho(u_1, u) \leq \epsilon} \left| Z_{g,k,m}(u_1) - Z_{g,k,m}(u) \right| > t/2 \right) \leq \epsilon^2,$$

and condition (iii) of Theorem 14 holds. The claim of the lemma now follows by applying Theorem 14.

**Q.E.D.**

**Proofs of Theorems**

**PROOF OF THEOREM 1:** The proof consists of two steps. First, we show that $\sqrt{G}(\hat{\beta}(u) - \tilde{\beta}(u)) = o_p(1)$ uniformly over $u \in U$ where $\tilde{\beta}(u)$ is defined in (19). Second, we show that $\sqrt{G}(\tilde{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot)$ in $\ell^\infty(U)$. Combining these steps gives the result.

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

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

By Lemma 1, $X'W/G \rightarrow_p Q_{xw}$ and $W'W/G \rightarrow_p Q_{ww}$ 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

$$\hat{S} = \left( \hat{Q}_{xw} \hat{Q}_{ww}^{-1} \hat{Q}_{xw}' \right)^{-1} \hat{Q}_{xw} \hat{Q}_{ww}^{-1} \rightarrow_p \left( Q_{xw} Q_{ww} Q_{xw}' \right)^{-1} Q_{xw} Q_{ww}^{-1} = S.$$ (34)

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

$$S(u) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left( \hat{\alpha}_g(u) - \alpha_g(u) \right) w'_g = o_p(1)$$

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

$$S_1(u) = - \frac{1}{\sqrt{G}} \sum_{g=1}^{G} J_g^{-1}(u) \mathbb{E} \left( h_{g}(u, u) \right) w'_g / \sqrt{N_g},$$

$$S_2(u) = \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \left( J_g^{-1}(u) \mathbb{E} \left( h_{g}(u, u) \right) + \sqrt{N_g} (\hat{\alpha}_g(u) - \alpha_g(u)) \right) w'_g / \sqrt{N_g}.$$
Since \( N_G = \min_{g=1, \ldots, G} N_g \to \infty \) by Assumption 3, Lemma 10 implies that \( S_1(u) = o_p(1) \) uniformly over \( u \in \mathcal{U} \).

Consider \( S_2(u) \). Let

\[
K_g = C \sqrt{N^{-1}_g \log N_g}
\]

for sufficiently large constant \( C > 0 \) so that Theorem 3 implies that

\[
P\left( \sup_{u \in \mathcal{U}} \| \hat{\alpha}_g(u) - \alpha_g(u) \| > K_g \right) \leq CN_g^{-3}.
\]

Let \( \mathcal{D}_G \) be the event that

\[
\sup_{u \in \mathcal{U}} \| \hat{\alpha}_g(u) - \alpha_g(u) \| \leq K_g, \quad \text{for all } g = 1, \ldots, G,
\]

and let \( \mathcal{D}_G^c \) be the event that \( \mathcal{D}_G \) does not hold. By the union bound, \( P(\mathcal{D}_G^c) \leq CN_g^{-3} \). 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, \( \|S_2(u)1\{\mathcal{D}_G\}\| \leq C \sum_{g=1}^G (r_{1,g} + r_{2,g} + r_{3,g})/\sqrt{GN_g} \)

where

\[
r_{1,g} = \sup_{u \in \mathcal{U}} \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq K_g} \| J_g^{-1}(u) \left( \mathbb{G}_g^g(h_{a,u}) - \mathbb{G}_g^g(h_{\hat{\alpha}_g(u),u}) \right) \| w_g \|
\]

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

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

\[\left. - (\alpha - \alpha_g(u))) \right\| w_g \|
\]

We bound the three terms \( r_{1,g}, r_{2,g}, \) and \( r_{3,g} \) in turn. By Lemma 4 and Hölder’s inequality,

\[
E[r_{1,g}] \leq (E[\|w_g\|^2])^{1/2}
\]

\[
\times \left( E\left[ \sup_{u \in \mathcal{U}} \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq K_g} \| \mathbb{G}_g^g(h_{a,u}) - \mathbb{G}_g^g(h_{\hat{\alpha}_g(u),u}) \|^2 \right] \right)^{1/2}
\]

\[
\leq C \left( \frac{\log N_g}{N_g} \log N_g \right)^{1/2} = \frac{(\log N_g)^{3/4}}{N_g^{1/4}},
\]
where the second line follows from the definition of \( K_g \), Assumption 2(iv), and Lemma 9. Further, using Lemma 4 again gives

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

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,

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

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

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

\[
\sqrt{G}(\hat{\beta}(\cdot) - \beta(\cdot)) = \hat{S} \cdot \frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g e_g(\cdot).
\]

As explained in Step 1, \( \hat{S} \rightarrow_p S \). Also, by Lemma 3,

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

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,

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

where \( \mathbb{G} \) is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function \( C(u_1, u_2) = SJ(u_1, u_2)S' \). Combining (36) with Step 1 gives the asserted claim and completes the proof of the theorem.

\textit{Q.E.D.}
Proof of Theorem 2: Equation (34) 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,

\[
\hat{\alpha}_{g,1}(u) - x'_g \hat{\beta}(u) = (\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)) - x'_g (\hat{\beta}(u) - \beta(u)) + \varepsilon_g(u)
\]

\[
= I_{1,g}(u) - I_{2,g}(u) + \varepsilon_g(u),
\]

where \( I_{1,g}(u) = \hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u) \) and \( I_{2,g}(u) = x'_g (\hat{\beta}(u) - \beta(u)) \). Further, we have

\[
\frac{1}{G} \sum_{g=1}^{G} \varepsilon_g(u_1)\varepsilon_g(u_2)w_gw'_g \to_p J(u_1, u_2)
\]

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

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

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

\[
\frac{1}{G} \sum_{g=1}^{G} I_{1,g}(u_1)I_{1,g}(u_2)w_gw'_g \leq \frac{K_G^2}{G} \sum_{g=1}^{G} \| w_g \|^2 + o_p(1)
\]

\[
\leq O_p(K_G^2) + o_p(1) = o_p(1)
\]

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

\[
\frac{1}{G} \sum_{g=1}^{G} I_{1,g}(u_1)\varepsilon_g(u_2)w_gw'_g \leq \frac{K_G}{G} \sum_{g=1}^{G} |\varepsilon_g(u_2)| \| w_g \|^2 + o_p(1)
\]

\[
\leq \frac{K_G}{G} \sum_{g=1}^{G} \sup_{u \in \mathcal{U}} |\varepsilon_g(u)| \| w_g \|^2 + o_p(1)
\]

\[
= o_p(1)
\]

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,

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

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

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

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

**Proof of Theorem 3:** Recall the definition of the function $f_{\eta, \alpha, u}$ in (20). 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

$$\inf_{u \in \mathcal{U}} \inf_{\eta \in \mathbb{R}^d : \| \eta \| = 1} \sum_{i=1}^{N_g} f_{\eta, \alpha_g(u) + x\eta, u}(z_{ig}, y_{ig}) / N_g > 0. \tag{37}$$

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

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

if the constant $\bar{c}$ in the statement of the theorem is sufficiently small. Therefore, it follows that (37) holds if

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

\begin{equation}
\inf_{u \in \mathcal{U}} \inf_{\eta, \alpha \in \mathbb{R}^{d_z}} \mathbb{G}^g(f_{\eta, \alpha, u}) \geq -c_2 \sqrt{N_g}.
\end{equation}

Note that for any \( \eta \in \mathbb{R}^{d_z} \) satisfying \( \| \eta \| = 1 \), \( |f_{\eta, \alpha, u}| \leq 2 \| z \| \leq C \) 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 (38) holds with probability not smaller than

\[ 1 - C \exp\left(-c_2^2 N_g\right) \]

for some \( c, C > 0 \). The asserted claim follows. \( Q.E.D. \)

PROOF OF THEOREM 4: Observe that the statement

\[ \beta_1(u) \not\in \left[ \hat{\beta}_1(u) - \hat{c}_{1-\alpha} \sqrt{\hat{V}(u) / G} \right] \]

for some \( u \in \mathcal{U} \) is equivalent to the statement that \( T > \hat{c}_{1-\alpha} \). Therefore, it suffices to prove that

\begin{equation}
P(T > \hat{c}_{1-\alpha}) \to \alpha.
\end{equation}

To prove (39), recall the process \( \mathbb{G}(\cdot) = (\mathbb{G}_1(u), \ldots, \mathbb{G}_{d_z}(u))' \) appearing in Theorem 1. Define a Gaussian process \( \tilde{\mathbb{G}}(\cdot) \) on \( \mathcal{U} \) with values in \( \mathbb{R} \) by

\[ \tilde{\mathbb{G}}(u) = V(u)^{-1/2} \mathbb{G}_1(u), \quad u \in \mathcal{U}, \]

where \( V(u) = C_{1,1}(u, u) \), the (1, 1)st component of \( \mathbb{C}(u, u) = SJ(u, u)S' \). 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 \( \tilde{\mathbb{G}}(\cdot) \) also has uniformly continuous sample paths. The covariance function of the process \( \tilde{\mathbb{G}}(\cdot) \) is

\[ \mathbb{C}(u_1, u_2) = V(u_1)^{-1/2} C_{1,1}(u_1, u_2) V(u_2)^{-1/2}. \]

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

\begin{align*}
\hat{\mathbb{G}}_G(u) &= \frac{1}{\sqrt{G \hat{V}(u)}} \sum_{g=1}^{G} \left( \epsilon_g (\hat{x}_g(u) - x'_g \hat{\beta}(u)) \hat{w}_{g,1}^s \right), \quad u \in \mathcal{U}, \\
\bar{\mathbb{G}}_G(u) &= \frac{1}{\sqrt{GV(u)}} \sum_{g=1}^{G} \epsilon_g \hat{\varepsilon}_g(u) w_{g,1}^s, \quad u \in \mathcal{U},
\end{align*}
where $w_{g,1}$ and $\hat{w}_{g,1}$ are the first component of the vectors $Sw_g$ and $\hat{Sw}_g$, respectively, and $\hat{V}(u) = \hat{C}_{1,1}(u, u)$.

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

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

$$
P\left(\sup_{u \in \mathcal{U}} |\hat{G}_G(u)| > \beta_{1-a, u, \mathcal{U}} - \eta\right) \leq \alpha + \delta,
$$

$$
P\left(\sup_{u \in \mathcal{U}} |\hat{G}_G(u)| > \beta_{1-a+\eta, u, \mathcal{U}} + \eta\right) \geq \alpha - \delta.
$$

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

$$
P(T > \beta_{1-a, u, \mathcal{U}} - \eta) \leq \alpha + \delta + o(1),
$$

$$
P(T > \beta_{1-a+\eta, u, \mathcal{U}} + \eta) \geq \alpha - \delta + o(1).
$$

Hence, to prove (39), it suffices to show that for any $\eta > 0$,

$$
P(\beta_{1-a, u, \mathcal{U}} - \eta \leq \hat{c}_{1-a} \leq \beta_{1-a+\eta, u, \mathcal{U}} + \eta) \to 1.
$$

To prove (40), fix some $\eta > 0$. Since $\hat{G}(\cdot)$ has uniformly continuous sample paths, there exists a finite $\mathcal{U}(\eta, 1) \subset \mathcal{U}$ such that

$$
\beta_{1-a, u, \mathcal{U}} - \eta \leq \beta_{1-a-\eta/2, \mathcal{U}(\eta, 1)} - \eta/2,
$$

$$
\beta_{1-a+\eta, u, \mathcal{U}} + \eta \geq \beta_{1-a+\eta/2, \mathcal{U}(\eta, 1)} + \eta/2.
$$

Further, let $\mathcal{A}_G$ be the event that $G^{-1} \sum_{g=1}^{G} (w_{g,1})^2 \leq C$ 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 $u_1, u_2 \in \mathcal{U},$

$$
E_x \left[ \left( \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \epsilon_g \rho_g(u_2) - \epsilon_g(u_1) w_{g,1} \right)^2 \right] = \frac{1}{G} \sum_{g=1}^{G} (\epsilon_g(u_2) - \epsilon_g(u_1))^2 (w_{g,1})^2 \leq C |u_2 - u_1|^2
$$

where $w_{g,1}$ and $\hat{w}_{g,1}$ are the first component of the vectors $Sw_g$ and $\hat{Sw}_g$, respectively, and $\hat{V}(u) = \hat{C}_{1,1}(u, u)$. Theorem 1 combined with the continuous mapping theorem implies $T \Rightarrow \sup_{u \in \mathcal{U}} |\hat{G}_G(u)|$, and so Theorem 2.1 of Chernozhukov, Chetverikov, and Kato (2014b) implies that $\sup_{u \in \mathcal{U}} |\hat{G}_G(u)|$ has continuous distribution. Therefore, for any $\delta > 0$, there exists $\eta > 0$ such that

$$
P\left(\sup_{u \in \mathcal{U}} |\hat{G}_G(u)| > \beta_{1-a, u, \mathcal{U}} - \eta\right) \leq \alpha + \delta,
$$

$$
P\left(\sup_{u \in \mathcal{U}} |\hat{G}_G(u)| > \beta_{1-a+\eta, u, \mathcal{U}} + \eta\right) \geq \alpha - \delta.
$$

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

$$
P(T > \beta_{1-a, u, \mathcal{U}} - \eta) \leq \alpha + \delta + o(1),
$$

$$
P(T > \beta_{1-a+\eta, u, \mathcal{U}} + \eta) \geq \alpha - \delta + o(1).
$$

Hence, to prove (39), it suffices to show that for any $\eta > 0$,

$$
P(\beta_{1-a, u, \mathcal{U}} - \eta \leq \hat{c}_{1-a} \leq \beta_{1-a+\eta, u, \mathcal{U}} + \eta) \to 1.
$$

To prove (40), fix some $\eta > 0$. Since $\hat{G}(\cdot)$ has uniformly continuous sample paths, there exists a finite $\mathcal{U}(\eta, 1) \subset \mathcal{U}$ such that

$$
\beta_{1-a, u, \mathcal{U}} - \eta \leq \beta_{1-a-\eta/2, \mathcal{U}(\eta, 1)} - \eta/2,
$$

$$
\beta_{1-a+\eta, u, \mathcal{U}} + \eta \geq \beta_{1-a+\eta/2, \mathcal{U}(\eta, 1)} + \eta/2.
$$

Further, let $\mathcal{A}_G$ be the event that $G^{-1} \sum_{g=1}^{G} (w_{g,1})^2 \leq C$ 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 $u_1, u_2 \in \mathcal{U},$

$$
E_x \left[ \left( \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \epsilon_g \rho_g(u_2) - \epsilon_g(u_1) w_{g,1} \right)^2 \right] = \frac{1}{G} \sum_{g=1}^{G} (\epsilon_g(u_2) - \epsilon_g(u_1))^2 (w_{g,1})^2 \leq C |u_2 - u_1|^2
$$
by Assumption 6(iii) where $E_{\cdot}[\cdot]$ denotes expectation with respect to the distribution of $\epsilon_1, \ldots, \epsilon_G$ (and keeping everything else fixed). Therefore, combining Borell’s inequality (see Proposition A.2.1 of Van der Vaart and Wellner (1996)) and Corollary 2.2.8 of Van der Vaart and Wellner (1996) shows that one can find finite $U(\eta, 2) \subset U$ such that, on $A_G$,

(43) $c_1 - a + \eta/2, U(\eta, 2), G + \eta/3 \geq c_1 - a + \eta/3, U(\eta, 2), G + \eta/4,$

(44) $c_1 - a - \eta/2, U(\eta, 2), G - \eta/3 \leq c_1 - a - \eta/3, U(\eta, 2), G - \eta/4.$

Now, observe that whenever the inequalities (41)–(44) are satisfied, the same inequalities are also satisfied with $U(\eta, 1)$ and $U(\eta, 2)$ replaced by $U(\eta) = U(\eta, 1) \cap U(\eta, 2)$.

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

$$
\tilde{C}_G(u_1, u_2) = V(u_1)^{-1/2} \left( \frac{1}{G} \sum_{g=1}^{G} e_g(u_1) e_g(u_2) (w_{g,1}^S)^2 \right) V(u_2)^{-1/2}.
$$

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

$$
P(c^0_{1 - a + \eta/2, U(\eta)} + \eta/2 > c_1 - a + \eta/2, U(\eta), G + \eta/3) \rightarrow 1,$$

$$
P(c^0_{1 - a - \eta/2, U(\eta)} - \eta/2 < c_1 - a - \eta/2, U(\eta), G - \eta/3) \rightarrow 1.$$

Combining this with inequalities (41)–(44) where we replace $U(\eta, 1)$ and $U(\eta, 2)$ by $U(\eta)$ gives

$$
P(c^0_{1 - a, U(\eta)} + \eta > c_1 - a + \eta/3, U(\eta), G + \eta/4) \rightarrow 1,$$

$$
P(c^0_{1 - a, U(\eta)} - \eta < c_1 - a - \eta/3, U(\eta), G - \eta/4) \rightarrow 1.$$

To complete the proof, it suffices to show that

(45) $P(c_1 - a - \eta/3, U(\eta), G - \eta/4 \leq \tilde{c}_1 - a \leq c_1 - a + \eta/3, U(\eta), G + \eta/4) \rightarrow 1.$

To prove (45), observe that

$$
\sup_{u \in U(\eta)} \left| \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \epsilon_g X_g'(\hat{\beta}(u) - \beta(u)) w_{g,1}^S \right| \leq \sup_{u \in U(\eta)} \| \hat{\beta}(u) - \beta(u) \| \cdot \left\| \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \epsilon_g w_{g,1}^S X_g \right\| \rightarrow P 0
$$
since \( \sup_{u \in \mathcal{U}} \| \beta(u) - \beta(u) \| \to 0 \) by Theorem 1 and \( \| G^{-1/2} \sum_{g=1}^{G} \varepsilon_g w_{g,1} x_g \| = O_P(1) \) by Assumptions 2(iv) and 4(i). Also,

\[
\sup_{u \in \mathcal{U}} \left\| \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \varepsilon_g (\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)) w_{g,1} \right\| \to 0
\]

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

\[
\sup_{u \in \mathcal{U}} \left\| \tilde{G}_G(u) - \tilde{G}_G(u) \right\| \to 0.
\]

Since \( \hat{c}_{1-a} \) is the \((1 - \alpha)\) conditional quantile of \( \sup_{u \in \mathcal{U}} | \tilde{G}(u) | \) given the data and \( c_{\beta,\mathcal{U},G} \) is the \( \beta \)th conditional quantile of \( \sup_{u \in \mathcal{U}} | \tilde{G}(u) | \) given the data, (45) follows. This completes the proof of the theorem. Q.E.D.

**Proof of Theorem 5:** We split the proof into two steps.

**Step 1.** Here we wish to show that for sufficiently large \( C > 0 \),

\[
P \left( \max_{1 \leq g \leq G} \left\| J^{-1}_g(u) \mathbb{G}^x (h_{\alpha_g(u),u}) + \sqrt{N_g} (\hat{\alpha}_g - \alpha_g) \right\| > \frac{C (\log N G)^{3/4}}{N^{1/4} G} \right) \to 0.
\]

Set \( K_g = C (N_g^{-1} \log N_g)^{1/2} \) for sufficiently large \( C > 0 \) so that Theorem 3 implies that

\[
P \left( \left\| \hat{\alpha}_g(u) - \alpha_g(u) \right\| > K_g \right) \leq C N_g^{-3}.
\]

Let \( D_G \) be the event that

\[
\left\| \hat{\alpha}_g(u) - \alpha_g(u) \right\| \leq K_g, \quad \text{for all } g = 1, \ldots, G,
\]

and let \( D_G^c \) be the event that \( D_G \) does not hold. By the union bound, \( P(D_G^c) \leq C G N_g^{-3} \to 0 \).

Now, on the event \( D_G \),

\[
\left\| J^{-1}_g(u) \mathbb{G}^x (h_{\alpha_g(u),u}) + \sqrt{N_g} (\hat{\alpha}_g - \alpha_g) \right\| \leq r_{1,g} + r_{2,g} + r_{3,g},
\]

where

\[
r_{1,g} = \sup_{a \in \mathbb{R}^d : \| a - \alpha_g(u) \| \leq K_g} \left\| J^{-1}_g(u) \left( \mathbb{G}^x (h_{a,u}) - \mathbb{G}^x (h_{\alpha_g(u),u}) \right) \right\|,
\]

\[
r_{2,g} = \sup_{a \in \mathbb{R}^d : \| a - \alpha_g(u) \| \leq K_g} \left\| \sqrt{N_g} (\hat{\alpha}_g - \alpha_g) \right\|
\]

\[
r_{3,g} = \sup_{a \in \mathbb{R}^d : \| a - \alpha_g(u) \| \leq K_g} \left\| J^{-1}_g(u) \mathbb{G}^x (h_{a,u}) \right\|
\]
\[ r_{2,g} = \left\| J_g^{-1}(u) \frac{1}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g,u}(z_{ig}, y_{ig}) \right\|, \]
\[ r_{3,g} = \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq K_g} \left\| E_g \left[ \sqrt{N_g} (J_g^{-1}(u) h_{\alpha,u}(z_{ig}, y_{ig}) - (\alpha - \alpha_g(u))) \right] \right\|. \]

By Lemma 4 and optimality of \( \hat{\alpha}_g(u) \),
\[ r_{2,g} \leq \frac{C}{\sqrt{N_g}} \sum_{i=1}^{N_g} h_{\hat{\alpha}_g,u}(z_{ig}, y_{ig}) \leq \frac{C}{\sqrt{N_g}}. \]

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, e.g., Theorem B.1 in Chernozhukov, Chetverikov, and Kato (2014b)),
\[ r_{1,g} \leq C \sup_{\alpha \in \mathbb{R}^d : \|\alpha - \alpha_g(u)\| \leq K_g} \left\| G^g(h_{\alpha,u}) - G^g(h_{\alpha_g(u),u}) \right\| \leq C \sqrt{K_g} \log G = \frac{C \log^{3/4} N_g}{N_g^{1/4}}, \]
with probability at least \( 1 - G^{-2} \). Combining these bounds gives (46) and completes this step.

**Step 2.** Here we complete the proof. For \( g = 1, \ldots, G \) and \( i = 1, \ldots, \bar{N}_g \), define \( q_{ig} \) as follows. If \( i > N_g \), set \( q_{ig} = 0 \). If \( i \leq N_g \), set
\[ q_{ig} = (\bar{N}_G/N_g)^{1/2} I_g^{-1/2} \tilde{z}_{ig} \left( \left\{ y_{ig} \leq \tilde{z}_{ig}' \alpha_g(u) \right\} - u \right), \]
where \( \tilde{z}_{ig} \) 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
\[ P \left( \max_{1 \leq g \leq G} \sqrt{N_g/I_g} \left| \hat{\alpha}_{g,1} - \alpha_{g,1} \right| \leq c_{1-\alpha}^M \right) \leq P \left( \max_{1 \leq g \leq G} \left\| \frac{1}{\sqrt{N_g}} \sum_{g=1}^{\bar{N}_G} (q_{ig} - E_g[q_{ig}]) \right\| \leq c_{1-\alpha}^M + \frac{C \log^{3/4} N_g}{N_g^{1/4}} \right) + o(1). \]
In turn, since, under our assumptions, \(|q_{i\ell}| \leq C\), by Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014c), the probability in (47) is bounded from above by

\[
P\left( \max_{1 \leq g \leq G} |Y_{g}| \leq c_{1-\alpha}^{M} + \frac{C \log^{3/4} N_{G}}{N_{G}^{1/4}} \right) + o(1)
\]

\[
\leq P\left( \max_{1 \leq g \leq G} |Y_{g}| \leq c_{1-\alpha}^{M} \right) + \frac{C (\log^{3/4} N_{G}) \cdot (\log^{1/2} G)}{N_{G}^{1/4}} + o(1)
\]

\[= 1 - \alpha + o(1),\]

where in the second line we used Theorem 3 in Chernozhukov, Chetverikov, and Kato (2015). Thus,

\[
(48) \quad P\left( \max_{1 \leq g \leq G} \sqrt{\frac{N_{g}/I_{g}}{G}} |\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)| \leq c_{1-\alpha}^{M} \right)
\]

\[= 1 - \alpha + o(1).\]

Similar arguments also give

\[
(49) \quad P\left( \max_{1 \leq g \leq G} \sqrt{\frac{N_{g}/I_{g}}{G}} |\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)| \leq c_{1-\alpha}^{M} \right)
\]

\[\geq 1 - \alpha - o(1).\]

Rearranging the terms under the probability signs in (48) and (49) completes the proof of the theorem. \(Q.E.D.\)

APPENDIX G: PROOFS OF THEOREMS 6–8

The proofs are analogous to those of Theorems 1, 2, and 4. Therefore, we only discuss important differences. First, the constants \(c, C > 0\) in the proofs now depend on \(c_{M}, c_{f}, C_{M}, C_{f}, C_{L}, \text{and } \hat{C}\). 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 \(k = 1, \ldots, d_{x}\) and \(l = 1, \ldots, d_{w}\),

\[
E\left[ \left( \frac{1}{G} \sum_{g=1}^{G} (x_{g,k} w_{g,l} - E[x_{g,k} w_{g,l}]) \right)^{2} \right]
\]

\[= \frac{1}{G^{2}} \sum_{m=1}^{M} E\left[ \left( \sum_{g \in C_{G}(m)} (x_{g,k} w_{g,l} - E[x_{g,k} w_{g,l}]) \right)^{2} \right]\]
\[ \leq \frac{C}{G^2} \sum_{m=1}^{M} \mathbb{E}\left[ \sum_{g \in C_G(m)} (x_{g,k} w_{g,l} - E[x_{g,k} w_{g,l}])^2 \right] \]

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

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 \( G \to \infty \),

\[ \frac{1}{G} \sum_{m=1}^{M} \left( \sum_{g \in C_G(m)} \epsilon_g(u_1)w_g \right) \left( \sum_{g \in C_G(m)} \epsilon_g(u_1)w_g' \right) \to p JCS(u_1, u_2) \]

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

\[ \frac{1}{G} \sum_{m=1}^{M} \mathbb{E}\left[ \left( \sum_{g \in C_G(m)} \epsilon_g(u_1)w_g \right) \left( \sum_{g \in C_G(m)} \epsilon_g(u_2)w_g' \right) \right] \to JCS(u_1, u_2) \]

uniformly over \( u_1, u_2 \in \mathcal{U} \). Further, for \( \delta = cM/4 \) and \( k, l = 1, \ldots, d_w \),

\[ \mathbb{E}\left[ \left| \left( \sum_{g \in C_G(m)} \epsilon_g(u_1)w_{g,k} \right) \left( \sum_{g \in C_G(m)} \epsilon_g(u_2)w_{g,l} \right) \right|^{1+\delta} \right] \leq C \mathbb{E}\left[ \sum_{g, g' \in C_G(m)} \left| \epsilon_g(u_1)w_{g,k} \epsilon_{g'}(u_2)w_{g',l} \right|^{1+\delta} \right] \leq C \mathbb{E}\left[ \sum_{g, g' \in C_G(m)} \left( \left| \epsilon_g(u_1)w_{g,k} \right|^{2+2\delta} + \left| \epsilon_{g'}(u_2)w_{g',l} \right|^{2+2\delta} \right) \right] \leq C, \]

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 (50) is analogous to the proof used in Lemma 2.

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

\[ \left( \frac{1}{\sqrt{G}} \sum_{g=1}^{G} w_g \epsilon_g(u) \right)_{u \in \mathcal{U}'} \Rightarrow (N(u))_{u \in \mathcal{U}'}, \]

where \( (N(u))_{u \in \mathcal{U}'} \) is a zero-mean Gaussian vector with covariance function \( JCS(u_1, u_2) \) 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 \( Z_g(u) = G^{-1/2}w_{g,k}e_g(u) \) by \( Z_m(u) = G^{-1/2} \sum_{g \in C_G(m)} w_{g,k}e_g(u) \), and we replace sums over \( g = 1, \ldots, G \) by sums over \( m = 1, \ldots, M \) where appropriate.

The statement of Lemma 10 holds without changes, but in the proof, we replace \( Z_g(u) = v_{g,k}(u)w_g/\sqrt{G} \) by \( Z_m(u) = \sum_{g \in C_G(m)} v_{g,k}(u)w_g/\sqrt{G} \) and we replace sums over \( g = 1, \ldots, G \) by sums over \( m = 1, \ldots, M \) 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

\[
\left\| \frac{1}{G} \sum_{m=1}^{M} \left( \sum_{g \in C_G(m)} I_{1,g}(u_1)w_g \right) \left( \sum_{g \in C_G(m)} I_{1,g}(u_2)w_g' \right) \right\| \leq \frac{1}{G} \sum_{m=1}^{M} \sum_{g, g' \in C_G(m)} \left\| I_{1,g}(u_1)I_{1,g'}(u_2)w_g w_g' \right\| \\
\leq \frac{K^2}{G} \sum_{g=1}^{G} \| w_g \|^2 + o_P(1) = o_P(1),
\]

and we bound all other terms in the proof similarly. The proof of Theorem 8 is analogous to that of Theorem 4.

**APPENDIX H: TOOLS**

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

Let \((T, \rho)\) be a semi-metric space. For \( \varepsilon > 0 \), an \( \varepsilon \)-net of \((T, \rho)\) is a subset \( T_\varepsilon \) 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 \( \varepsilon \)-covering number \( N(\varepsilon, T, \rho) \) of \( T \) is the infimum of the cardinality of \( \varepsilon \)-nets of \( T \), that is, \( N(\varepsilon, T, \rho) = \inf \{ \text{Card}(T_\varepsilon) : T_\varepsilon \text{ is an } \varepsilon \text{ net of } T \} \).

Let \( F \) be a class of measurable functions defined on some measurable space \((S, S)\). For any probability measure \( Q \) on \((S, S)\) and \( p \geq 1 \), 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 \( F \) is called VC subgraph class if the collection of all subgraphs of the functions in \( F \) 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 \( F \) is VC type class of functions with an envelope \( F : S \to \mathbb{R}_+ \) and constants \( A \geq \varepsilon \), and \( v \geq 1 \) if all functions in \( F \) are bounded in absolute value by \( F \) and the following condition holds:

\[
\sup_{Q} N(\varepsilon\|F\|_{Q,2}, F, L_2(Q)) \leq (A/\varepsilon)^v
\]
for all $\varepsilon \in (0, 1)$ where the supremum is taken over all finitely discrete probability measures $Q$ on $(S, \mathcal{S})$.

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

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

The following theorems are used in Appendix F:

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

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

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).  

**Q.E.D.**

**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 $f = (f_1, \ldots, f_k)$, $g = (g_1, \ldots, g_k) \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k = \mathcal{F}$ and every $s \in S$, where $L_1, \ldots, L_k$ are nonnegative 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

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

for every $0 < \varepsilon < 1$.


**Q.E.D.**
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 $\|M\|_2 = (E[M^2])^{1/2}$.


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

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

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.


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

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

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). Q.E.D.

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

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

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

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

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

**THEOREM 14**—Bracketing by Gaussian Hypotheses: For each $n$, let $Z_{n1}, \ldots, Z_{nm}$ be independent stochastic processes indexed by an arbitrary index set $F$. Suppose that there exists a Gaussian-dominated semi-metric $\rho$ on $F$ such that

(i) $\sum_{i=1}^{mn} E[\|Z_{ni}\|_F \cdot 1\{\|Z_{ni}\|_F > \eta\}] \to 0$, for every $\eta > 0$,

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

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

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

**PROOF:** See Theorem 2.11.11 in Van der Vaart and Wellner (1996).

Q.E.D.

**ADDITIONAL REFERENCES**


[11]

Dept. of Economics, UCLA, 315 Portola Plaza, Los Angeles, CA 90095, U.S.A.;
chetverikov@econ.ucla.edu,
Dept. of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305,
U.S.A. and NBER; bjlarsen@stanford.edu,
and
Haas School of Business, University of California at Berkeley, Berkeley, CA
94720, U.S.A.; cjpalmer@berkeley.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received December, 2013; final revision received August, 2015.