*Econometrica Supplementary Material*

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

(*Econometrica*, Vol. 84, No. 2, March 2016, 809–833)

BY DENIS CHETVERIKOV, BRADLEY LARSEN, AND CHRISTOPHER PALMER

## 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:

$$y_{ig} = z_{ig}\gamma(u_{ig}) + \delta(u) + x_g\beta(u_{ig}) + \varepsilon_g(u_{ig}), \tag{10}$$

$$x_g = \pi w_g + \eta_g + \nu_g, \tag{11}$$

$$\varepsilon_g(u) = u\eta_g - \frac{u}{2}, \tag{12}$$

where $w_g$, $\nu_g$, and $z_{ig}$ are each distributed $\exp(0.25^*N[0,1])$; $u_{ig}$ 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 $\varepsilon_g(u) = u\eta_g - \frac{u}{2}$ implies $E[\varepsilon_g(u)|w_g] = E[u\eta_g - u/2|w_g] = E[u\eta_g - u/2] = u/2 - u/2 = 0$. The quantile coefficient functions are $\gamma(u) = \beta(u) = u^{1/2}$ and $\delta(u) = u/2$. The parameter $\pi = 1$.

We employ three variants of the data generating process described in (10)–(12). The first case is exactly as in (10)–(12), with the group-level treatment of interest, $x_g$, being endogenous (correlated with $\varepsilon_g$ 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 $\varepsilon_g$). 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 $\varepsilon_g$. In the third case, $x_g$ is exogenous *and* no group-level unobservables are included, where we set $x_g = w_g$ and $\varepsilon_g = 0$. In this latter case, both grouped quantile regression and standard quantile regression should be consistent.

We perform these exercises with the number of groups ($G$) and the number of observations per group ($N$) given by $(N, G) = (25, 25), (200, 25), (25, 200), (200, 200)$. 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 ($u = 0.1, \dots, 0.9$) 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 $N$

 DOI: 10.3982/ECTA12121

TABLE A.I

BIAS OF GROUPED IV QUANTILE REGRESSION VERSUS STANDARD QUANTILE REGRESSION[a]

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

[a]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 $u \in \{0.1, \dots, 0.9\}$ 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).

## 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* $g = 1, \dots, G$ *and* $x \in (0, \bar{c})$,

$$P\Big(\sup_{u\in\mathcal{U}}\big\|\hat{\alpha}_g(u) - \alpha_g(u)\big\| > x\Big) \le Ce^{-cx^2N_g}. \tag{13}$$

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 $\hat{\alpha}_{g,1}(u)$ 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 $\mathcal{U}$. Observe that $\beta(u)$ is a $d_x$-vector, that is, $\beta(u) = (\beta_1(u), \dots, \beta_{d_x}(u))'$. Without loss of generality, we focus on $\beta_1(u)$, the first component of $\beta(u)$. Let $\hat{\beta}_1(u)$, $V(u)$, and $\hat{V}(u)$ denote the first component of $\hat{\beta}(u)$, the $(1,1)$ component of $\mathcal{C}(u,u)$, and the $(1,1)$ component of $\hat{\mathcal{C}}(u,u)$, respectively. Define

$$T = \sup_{u\in\mathcal{U}}\sqrt{G}\big|\hat{V}(u)^{-1/2}\big(\hat{\beta}_1(u) - \beta_1(u)\big)\big|, \tag{14}$$

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

$$\left[\hat{\beta}_1(u) - c_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}, \hat{\beta}_1(u) + c_{1-\alpha}\sqrt{\frac{\hat{V}(u)}{G}}\right]. \tag{15}$$

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

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

The multiplier bootstrap critical value $\hat{c}_{1-\alpha}$ is the conditional $(1-\alpha)$ quantile of $T^{\mathrm{MB}}$ 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*

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

*as* $G \to \infty$.

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 $[\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}]$ for $\alpha_{g,1}(u)$ that are adjusted for multiplicity of the effects, that is, we would like to have the bands satisfying

$$P\left(\alpha_{g,1}(u) \in \left[\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}\right] \text{ for all } g = 1, \ldots, G\right) \to 1-\alpha. \tag{16}$$

Thus, the confidence bands $[\hat{\alpha}^l_{g,1}, \hat{\alpha}^r_{g,1}]$ cover the true group-specific effects $\alpha_{g,1}$ for all $g = 1, \ldots, G$ 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 $G^{2/3}(\log N_G)/N_G \to 0$ as $G \to \infty$ 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 $u(1-u)J_g(u)^{-1}E_g[z_{ig}z_{ig}']J_g(u)^{-1}$; see, for example, Koenker (2005). Therefore, letting $c_{1-\alpha}$ be the $(1-\alpha)$ quantile of $|Y|$ where $Y \sim N(0,1)$, we obtain

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

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 $\alpha_{g,1}(u)$, $g = 1, \dots, G$. This is especially important given that $G$ is large. To fix this problem, we would like to adjust the constant $c_{1-\alpha}$ in (17) so that the events under the probability sign in (17) hold simultaneously for all $g = 1, \dots, G$ with probability asymptotically equal to $1-\alpha$. The theorem below shows that this can be achieved by replacing $c_{1-\alpha}$ with $c_{1-\alpha}^M$, the $(1-\alpha)$ quantile of $\max_{1\le g\le G}|Y_g|$ where $Y_1, \dots, Y_G$ are i.i.d. $N(0,1)$ random variables. To decrease technicalities, we assume in the theorem that all $I_g$'s are known.

THEOREM 5—Joint Inference on Group-Specific Effects: *Let Assumptions* 1–8 *hold. In addition, suppose that* $I_g \ge c_M$ *for all* $g = 1, \dots, G$ *and* $\bar{N}_G/N_G \le C_M$ *where* $N_G = \min_{1\le g\le G} N_g$ *and* $\bar{N}_G = \max_{1\le g\le G} N_g$. *Let* $c_{1-\alpha}^M$ *be the* $(1-\alpha)$ *quantile of* $\max_{1\le g\le G}|Y_g|$ *where* $Y_1, \dots, Y_G$ *are i.i.d.* $N(0,1)$ *random variables. Then*

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

*as* $G \to \infty$.

REMARK 5: We note that the size of the bands in this theorem, $\max_{1\le g\le G} 2 \times c_{1-\alpha}^M (I_g/N_g)^{1/2}$, is shrinking to zero as $G$ gets large. Indeed, under our assumptions, $\max_{1\le g\le G} I_g \le C$ for some constant $C$, which is independent of $G$. In addition, $c_{1-\alpha}^M \le (C\log G)^{1/2}$ for some absolute constant $C$. Therefore, $\max_{1\le g\le G} c_{1-\alpha}^M (I_g/N_g)^{1/2} \le (C\log G/N_G)^{1/2} \to 0$ 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 $\mathbb{C}_G : \{1, \dots, M\} \rightrightarrows \{1, \dots, G\}$ such that (i) for each $m = 1, \dots, M$, $\mathbb{C}_G(m)$ denotes the set of groups corresponding to cluster $m$, (ii) for $m, m' = 1, \dots, M$ with $m \neq m'$, the set $\mathbb{C}_G(m) \cap \mathbb{C}_G(m')$ is empty, and (iii) for any $g = 1, \dots, G$, there exists $m = 1, \dots, M$ such that $g \in \mathbb{C}_G(m)$. Thus, the correspondence $\mathbb{C}_G(\cdot)$ partitions groups into $M$ clusters. Using this notation, we replace Assumption 1 with the following condition:

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

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

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

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

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

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

Assumptions 6′(i) and 6′(iii) are the same as Assumptions 6(i) and 6(iii). Assumption 6′(ii) is a modification of Assumption 6(ii) adjusting the asymptotic covariance function of $G^{-1/2}\sum_{g=1}^{G}\varepsilon_g(\cdot)w_g$ to allow for clustering. When $\mathbb{C}_G(m)$ contains only one group for each $m = 1, \dots, M$, Assumption 6′(ii) reduces to Assumption 6(ii).

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

THEOREM 6—Asymptotic Distribution Under Cluster Sampling: *Let Assumptions* 1′, 2–5, 6′, 7, *and* 8 *hold*. *Then*

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

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

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

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

where

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

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

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

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

$\ldots, \beta_{d_x}(u))'$. As 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 $\hat{\beta}_1(u)$, $V^{\mathrm{CS}}(u)$, and $\hat{V}^{\mathrm{CS}}(u)$ denote the 1st component of $\hat{\beta}(u)$, the $(1,1)$st component of $\mathcal{C}^{\mathrm{CS}}(u,u)$, and the $(1,1)$st component of $\hat{\mathcal{C}}^{\mathrm{CS}}(u,u)$, respectively. Define

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

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

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

The multiplier bootstrap critical value $\hat{c}_{1-\alpha}$ is the conditional $(1-\alpha)$ quantile of $T^{\mathrm{MB}}$ 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 $J^{\mathrm{CS}}(u,u)$ are bounded away from zero uniformly over $u \in \mathcal{U}$*. *Then*

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

*as $G \to \infty$*.

## APPENDIX F: 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

$$A(u) = \big(\alpha_{1,1}(u), \dots, \alpha_{G,1}(u)\big)', \tag{19}$$
$$\tilde{\beta}(u) = \big(X'P_W X\big)^{-1}\big(X'P_W A(u)\big),$$
$$J_g(u) = E_g\big[z_{1g}z'_{1g}f_g\big(z'_{1g}\alpha_g(u)\big)\big].$$

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

$$f_{\eta,\alpha,u}(z, y) = \big(z'\eta\big) \cdot \big(1\{y \le z'\alpha\} - u\big). \tag{20}$$

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 $h_{\alpha,u} : \mathbb{R}^{d_z} \times \mathbb{R} \to \mathbb{R}^{d_z}$ be defined by

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

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 $k = 1, \dots, d_z$.

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

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

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

*Preliminary Lemmas*

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

LEMMA 1: *As* $G \to \infty$,

$$\hat{Q}_{xw} = \frac{1}{G}\sum_{g=1}^{G} x_g w'_g \to_p Q_{xw}, \tag{21}$$

$$\hat{Q}_{ww} = \frac{1}{G}\sum_{g=1}^{G} w_g w'_g \to_p Q_{ww}, \tag{22}$$

*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

$$\frac{1}{G}\sum_{g=1}^{G}\left(x_g w_g' - E\left[x_g w_g'\right]\right) \to_p 0. \tag{23}$$

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* $G \to \infty$,

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

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

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

Since

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

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

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

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

To this end, fix $u_1, u_2 \in \mathcal{U}$ and $k, l = 1, \dots, d_w$. We first show (24) for these values of $u_1$, $u_2$, $k$, and $l$. Note that we cannot use Chebyshev's inequality here because $E[(\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l})^2]$ is not necessarily finite. Instead, we use a more delicate method as presented in Theorem 2.1.7 of Tao (2012). Let $\delta = c_M/4$. Then by Hölder's inequality,

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

In turn,

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

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

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

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

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

$$E\big[|X_g|^{1+\delta}\big] \le C. \tag{25}$$

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 $X_{g,\le N} = X_g \cdot 1\{|X_g| \le N\}$ and $X_{g,>N} = X_g \cdot 1\{|X_g| > N\}$. Then by Fubini's theorem and Markov's inequality,

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

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

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

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

$$\begin{aligned}
&P\left(\left|\frac{1}{G}\sum_{g=1}^G X_{g,\le N}\right| > \varepsilon + CN^{-\delta}\right) \\
&\quad\le P\left(\left|\frac{1}{G}\sum_{g=1}^G \big(X_{g,\le N} - E[X_{g,\le N}]\big)\right| > \varepsilon\right)
\end{aligned}$$

$$\leq \frac{1}{\varepsilon^2 G^2} \sum_{g=1}^{G} E\big[X^2_{g,\leq N}\big]$$

$$\leq \frac{N^2}{\varepsilon^2 G}.$$

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

$$\sup_{u_1,u_2\in\mathcal{U}} \left| \frac{1}{G} \sum_{g=1}^{G} \big(\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l} - E\big[\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}\big]\big) \right|$$

$$\leq \max_{u_1,u_2\in\mathcal{U}_\delta} \left| \frac{1}{G} \sum_{g=1}^{G} \big(\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l} - E\big[\varepsilon_g(u_1)\varepsilon_g(u_2) w_{g,k} w_{g,l}\big]\big) \right|$$

$$+ \frac{2\delta}{G} \sum_{g=1}^{G} \Big(\sup_{u\in\mathcal{U}} \big|\varepsilon_g(u)\big| \cdot |w_{g,k} w_{g,l}| + E\Big[\sup_{u\in\mathcal{U}} \big|\varepsilon_g(u)\big| \cdot |w_{g,k} w_{g,l}|\Big]\Big)$$

$$= o_p(1) + \delta \cdot O_p(1)$$

by the result above and Chebyshev's inequality. Since $\delta$ is arbitrary, this completes the proof. *Q.E.D.*

LEMMA 3: *As* $G \to \infty$,

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

*where* $\mathbb{G}^0$ *is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function* $J(u_1, u_2)$ *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

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

where $(N(u))_{u\in\mathcal{U}'}$ is a zero-mean Gaussian vector with covariance function $J(u_1, u_2)$ for all $u_1, u_2 \in \mathcal{U}'$. Therefore, to prove the asserted claim, we can apply Theorem 14. In particular, it suffices to verify conditions (i)–(iii) of Theorem 14 with $Z_g(u) = G^{-1/2} w_{g,k}\varepsilon_g(u)$, $g = 1, \dots, G$ and $u \in \mathcal{U}$, for all $k = 1, \dots, d_w$. In the verification, we will use the Gaussian-dominated semi-metric $\rho : \mathcal{U} \times \mathcal{U} \to \mathbb{R}_+$ defined by $\rho(u_1, u_2) = C|u_2 - u_1|$ for sufficiently large constant $C > 0$; see discussion in front of Theorem 14 for the definition of Gaussian-dominated semi-metrics.

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

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

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

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

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

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

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

$$
\begin{aligned}
&\sup_{t>0} \sum_{g=1}^{G} t^2 P\Big(\sup_{\rho(u_1,u_2)\le 2\epsilon} \big|Z_g(u_2) - Z_g(u_1)\big| > t\Big) \\
&\quad \le \frac{1}{G}\sum_{g=1}^{G} E\Big[\sup_{\rho(u_1,u_2)\le 2\epsilon} \big|w_{g,k}\varepsilon_g(u_2) - w_{g,k}\varepsilon(u_1)\big|^2\Big] \\
&\quad \le C \sup_{\rho(u_1,u_2)\le 2\epsilon} |u_2 - u_1|^2 \le \epsilon^2
\end{aligned}
$$

by Assumptions 2(iv) and 6(iii) since the constant $C$ in the definition of $\rho(u_1, u_2)$ is large enough. 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* $g = 1, \dots, G$, *all eigenvalues of* $J_g(u)$ *are bounded from below by* $c$, *and* (ii) *for all* $u_1, u_2 \in \mathcal{U}$ *and* $g = 1, \dots, G$, $\|J_g^{-1}(u_2) - J_g^{-1}(u_1)\| \le C|u_2 - u_1|$.

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

$$\alpha' J_g(u)\alpha \ge c_f \alpha' E_g\big[z_{1g} z_{1g}'\big]\alpha \ge c_f c_M, \tag{26}$$

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

$$\begin{aligned}\big\|J_g^{-1}(u_2) - J_g^{-1}(u_1)\big\| &\le \big\|J_g^{-1}(u_1)\big\|\big\|J_g^{-1}(u_2)\big\|\big\|J_g(u_2) - J_g(u_1)\big\| \\ &\le \frac{\big\|J_g(u_2) - J_g(u_1)\big\|}{(c_f c_M)^2},\end{aligned}$$

where the second inequality follows from (26). Hence, it suffices to show that $\|J_g(u_2) - J_g(u_1)\| \le C|u_2 - u_1|$ for some $C > 0$. To this end, note that

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

where the second inequality follows from Assumptions 4(i) and 5.

Thus, if $|u_2 - u_1| < c_f/(C_M C_L)$, then $z_{1g}'\alpha_g(u_2) \in B_g(u_1, c_f)$, and so

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

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

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

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

LEMMA 5: *There exist constants* $c, C > 0$ *such that, for all* $g = 1, \dots, G$,

$$\big\| E_g\big[h_{\alpha,u}(z_{1g}, y_{1g})\big] - J_g(u)\big(\alpha - \alpha_g(u)\big)\big\| \le C\big\|\alpha - \alpha_g(u)\big\|^2, \tag{27}$$

$$E_g\big[\big(\alpha - \alpha_g(u)\big)' h_{\alpha,u}(z_{1g}, y_{1g})\big] \ge c\big\|\alpha - \alpha_g(u)\big\|^2, \tag{28}$$

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

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

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

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 $\|r_n(u)\| \le C\|\alpha - \alpha_g(u)\|^2$ 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 $\|(\alpha - \alpha_g(u))' r_n(u)\| \le c\|\alpha - \alpha_g(u)\|^2$ for an arbitrarily small constant $c > 0$. On the other hand,

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

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* $k = 1, \dots, d_z$, $\mathcal{H}_k$ *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 $\{x \in \mathbb{R}^{d_z+1} : a'x \le 0\}$ with $a$ varying over $\mathbb{R}^{d_z+1}$. It is well known that this is a VC subgraph class of sets; see, for example, exercise 14 of Chapter 2.6 in Van der Vaart and Wellner (1996). Further, note that

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

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

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

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

PROOF: Observe that

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

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

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

The asserted claim follows. *Q.E.D.*

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

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

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

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

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

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

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

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

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

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

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

and so Theorem 13 shows that

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

since

$$\begin{aligned}
&E_g\Big[\max_{1\le i\le N_g} \sup_{u_2\in\mathcal{U}:|u_2-u_1|\le\epsilon} \big|N_g^{-1/2}\big(h_{k,\alpha_g(u_2),u_2}(z_{ig}, y_{ig}) \\
&\qquad - h_{k,\alpha_g(u_1),u_1}(z_{ig}, y_{ig})\big)\big|^4\Big] \\
&\le N_g^{-1} \max_{1\le i\le N_g} E_g\Big[\sup_{u_2\in\mathcal{U}:|u_2-u_1|\le\epsilon} \big|\big(h_{k,\alpha_g(u_2),u_2}(z_{ig}, y_{ig}) \\
&\qquad - h_{k,\alpha_g(u_1),u_1}(z_{ig}, y_{ig})\big)\big|^4\Big] \\
&\le N_g^{-1} E_g\big[F^4(z_{ij}, y_{ig})\big] \le C\epsilon.
\end{aligned}$$

The asserted claim follows. *Q.E.D.*

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

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

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

PROOF: Observe that

$$E_g\Big[\sup_{u\in\mathcal{U}} \sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon} \big\|\mathbb{G}^g(h_{\alpha,u}) - \mathbb{G}^g(h_{\alpha_g(u),u})\big\|^2\Big] \tag{29}$$

$$\le C\sum_{k=1}^{d_z} E_g\Big[\sup_{u\in\mathcal{U}} \sup_{\alpha\in\mathbb{R}^{d_z}:\|\alpha-\alpha_g(u)\|\le\epsilon} \big|\mathbb{G}^g(h_{k,\alpha,u}) - \mathbb{G}^g(h_{k,\alpha_g(u),u})\big|^2\Big]. \tag{30}$$

Consider the function class

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

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

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

and so Theorem 13 gives

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

The asserted claim follows. *Q.E.D.*

LEMMA 10: *Uniformly over $u \in \mathcal{U}$,*

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

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

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

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

$$\sum_{g=1}^{G} Z_{g,k,m}(u) = O_p(1) \quad \text{uniformly over } u \in \mathcal{U} \tag{31}$$

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 $a > 0$,

$$\sum_{g=1}^{G} E\Big[\sup_{u\in\mathcal{U}}\big|Z_{g,k,m}(u)\big| \cdot 1\Big\{\sup_{u\in\mathcal{U}}\big|Z_{g,k,m}(u)\big| > a\Big\}\Big] \tag{32}$$
$$\leq a^{-1}\sum_{g=1}^{G} E\Big[\sup_{u\in\mathcal{U}} Z^2_{g,k,m}(u) \cdot 1\Big\{\sup_{u\in\mathcal{U}}\big|Z_{g,k,m}(u)\big| > a\Big\}\Big]$$
$$\leq \frac{1}{aG}\sum_{g=1}^{G} E\Big[\sup_{u\in\mathcal{U}}\big(v_{g,k}(u)w_{g,m}\big)^2 \cdot 1\Big\{\sup_{u\in\mathcal{U}}\big|v_{g,k}(u)w_{g,m}\big| > \sqrt{G}a\Big\}\Big].$$

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

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

uniformly over $g = 1, \dots, G$ 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 $u_1, u_2 \in \mathcal{U}$,

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

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

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

$$\leq CE_g\big[\big\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2} - h_{\alpha_g(u_1),u_1})\big\|^4\big] + CE_g\big[\big\|\mathbb{G}^g(h_{\alpha_g(u_1),u_1})\big\|^4\big]\cdot|u_2-u_1|^4,$$

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

$$E_g\big[\big\|\mathbb{G}^g(h_{\alpha_g(u_2),u_2} - h_{\alpha_g(u_1),u_1})\big\|^4\big] \leq C|u_2-u_1| \quad \text{and} \qquad (33)$$
$$E_g\big[\big\|\mathbb{G}^g(h_{\alpha_g(u_1),u_1})\big\|^4\big] \leq C,$$

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

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

and so

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

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 $u_1 \in \mathcal{U}$,

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

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

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

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 $u \in \mathcal{U}$,

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

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

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 $u \in \mathcal{U}$ where $\tilde\beta(u)$ is defined in (19). Second, we show that $\sqrt{G}(\tilde\beta(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{U})$. Combining these steps gives the result.

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

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

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

$$\hat S = \big(\hat Q_{xw}\hat Q_{ww}^{-1}\hat Q'_{xw}\big)^{-1}\hat Q_{xw}\hat Q_{ww}^{-1} \to_p \big(Q_{xw}Q_{ww}Q'_{xw}\big)^{-1}Q_{xw}Q_{ww}^{-1} = S. \tag{34}$$

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

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

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

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

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

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

Consider $S_2(u)$. Let

$$K_g = C\sqrt{N_g^{-1}\log N_g} \tag{35}$$

for sufficiently large constant $C > 0$ so that Theorem 3 implies that

$$P\Big(\sup_{u\in\mathcal{U}}\big\|\hat{\alpha}_g(u) - \alpha_g(u)\big\| > K_g\Big) \le CN_g^{-3}.$$

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

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

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

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

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

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

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

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

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

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

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

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

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,

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

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

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

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

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

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,

$$\sqrt{G}\big(\tilde{\beta}(\cdot)-\beta(\cdot)\big) \Rightarrow \mathbb{G}(\cdot), \quad \text{in} \quad \ell^\infty(\mathcal{U}), \tag{36}$$

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

*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,

$$\begin{aligned}\hat{\alpha}_{g,1}(u) - x_g'\hat{\beta}(u) &= \big(\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)\big) - x_g'\big(\hat{\beta}(u) - \beta(u)\big) + \varepsilon_g(u) \\ &= I_{1,g}(u) - I_{2,g}(u) + \varepsilon_g(u),\end{aligned}$$

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

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

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

$$P\Big(\max_{g=1,\dots,G} \sup_{u\in\mathcal{U}} \big\|\hat{\alpha}_g(u) - \alpha_g(u)\big\| > K_g\Big) \le CGN_g^{-3} = o(1)$$

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

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

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

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

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

$$E\Big[\sup_{u\in\mathcal{U}}\big|\varepsilon_g(u)\big|\|w_g\|^2\Big] \le \Big(E\Big[\sup_{u\in\mathcal{U}}\big|\varepsilon_g(u)\big|^2\Big]\Big)^{1/2}\big(E\big[\|w_g\|^4\big]\big)^{1/2} \le C$$

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

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

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

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

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

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

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

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

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

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

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

which in turn follows if

$$\inf_{u\in\mathcal{U}}\inf_{\eta,\alpha\in\mathbb{R}^{d_z};\|\eta\|=1}\mathbb{G}^g(f_{\eta,\alpha,u})\geq -cx\sqrt{N_g}. \tag{38}$$

Note that for any $\eta\in\mathbb{R}^{d_z}$ satisfying $\|\eta\|=1$, $|f_{\eta,\alpha,u}|\leq 2\|z_{ig}\|\leq C$ for some $C>0$ by Assumption 4(i). In addition, it follows from Lemma 6 and Theorem 9 that the conditions of Theorem 12 hold for the function class $\{f_{\eta,\alpha,u}\in\mathcal{F}: u\in\mathcal{U};\ \eta,\alpha\in\mathbb{R}^{d_z};\ \|\eta\|=1\}$. Therefore, Theorem 12 shows that (38) holds with probability not smaller than

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

for some $c, C>0$. The asserted claim follows. *Q.E.D.*

PROOF OF THEOREM 4: Observe that the statement

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

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

$$P(T>\hat c_{1-\alpha})\to\alpha. \tag{39}$$

To prove (39), recall the process $\mathbb{G}(\cdot)=(\mathbb{G}_1(u),\dots,\mathbb{G}_{d_x}(u))'$ 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 $V(u)=\mathcal{C}_{1,1}(u,u)$, the $(1,1)$st component of $\mathcal{C}(u,u)=SJ(u,u)S'$. It follows from conditions of the theorem that $V(u)$ is bounded away from zero uniformly over $u\in\mathcal{U}$. Therefore, since $\mathbb{G}(\cdot)$ has uniformly continuous sample paths, the process $\tilde{\mathbb{G}}(\cdot)$ also has uniformly continuous sample paths. The covariance function of the process $\tilde{\mathbb{G}}(\cdot)$ is

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

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

$$\hat{\mathbb{G}}_G(u)=\frac{1}{\sqrt{G\hat V(u)}}\sum_{g=1}^G\big(\epsilon_g\big(\hat\alpha_{g,1}(u)-x_g'\hat\beta(u)\big)\hat w_{g,1}^S\big),\quad u\in\mathcal{U},$$

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

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

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

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

$$P\Big(\sup_{u\in\mathcal{U}}\big|\tilde{\mathbb{G}}(u)\big| > c^0_{1-\alpha-\eta,\mathcal{U}} - \eta\Big) \le \alpha + \delta,$$
$$P\Big(\sup_{u\in\mathcal{U}}\big|\tilde{\mathbb{G}}(u)\big| > c^0_{1-\alpha+\eta,\mathcal{U}} + \eta\Big) \ge \alpha - \delta.$$

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

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

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

$$P\big(c^0_{1-\alpha-\eta,\mathcal{U}} - \eta \le \hat{c}_{1-\alpha} \le c^0_{1-\alpha+\eta,\mathcal{U}} + \eta\big) \to 1. \tag{40}$$

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

$$c^0_{1-\alpha-\eta,\mathcal{U}} - \eta \le c^0_{1-\alpha-\eta/2,\mathcal{U}(\eta,1)} - \eta/2, \tag{41}$$
$$c^0_{1-\alpha+\eta,\mathcal{U}} + \eta \ge c^0_{1-\alpha+\eta/2,\mathcal{U}(\eta,1)} + \eta/2. \tag{42}$$

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

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

by Assumption 6(iii) where $E_\epsilon[\cdot]$ denotes expectation with respect to the distribution of $\epsilon_1, \ldots, \epsilon_G$ (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 $\mathcal{U}(\eta, 2) \subset \mathcal{U}$ such that, on $\mathcal{A}_G$,

$$c_{1-\alpha+\eta/2,\mathcal{U}(\eta,2),G} + \eta/3 \geq c_{1-\alpha+\eta/3,\mathcal{U},G} + \eta/4, \tag{43}$$

$$c_{1-\alpha-\eta/2,\mathcal{U}(\eta,2),G} - \eta/3 \leq c_{1-\alpha-\eta/3,\mathcal{U},G} - \eta/4. \tag{44}$$

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

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

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

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

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

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

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

To complete the proof, it suffices to show that

$$P(c_{1-\alpha-\eta/3,\mathcal{U},G} - \eta/4 \leq \hat{c}_{1-\alpha} \leq c_{1-\alpha+\eta/3,\mathcal{U}} + \eta/4) \to 1. \tag{45}$$

To prove (45), observe that

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

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

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

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

$$\sup_{u\in\mathcal{U}}\big\|\tilde{\mathbb{G}}_G(u) - \hat{\mathbb{G}}_G(u)\big\| \to_p 0.$$

Since $\hat{c}_{1-\alpha}$ is the $(1-\alpha)$ conditional quantile of $\sup_{u\in\mathcal{U}} |\hat{\mathbb{G}}(u)|$ given the data and $c_{\beta,\mathcal{U},G}$ is the $\beta$th conditional quantile of $\sup_{u\in\mathcal{U}} |\tilde{\mathbb{G}}(u)|$ 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 $C > 0$,

$$P\left(\max_{1\le g\le G}\big\|J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u}) + \sqrt{N_g}(\hat{\alpha}_g - \alpha_g)\big\| > \frac{C(\log N_G)^{3/4}}{N_G^{1/4}}\right) \to 0. \tag{46}$$

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

$$P\big(\big\|\hat{\alpha}_g(u) - \alpha_g(u)\big\| > K_g\big) \le CN_g^{-3}.$$

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

$$\big\|\hat{\alpha}_g(u) - \alpha_g(u)\big\| \le K_g, \quad \text{for all } g = 1, \dots, G,$$

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

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

$$\big\|J_g^{-1}(u)\mathbb{G}^g(h_{\alpha_g(u),u}) + \sqrt{N_g}(\hat{\alpha}_g - \alpha_g)\big\| \le r_{1,g} + r_{2,g} + r_{3,g},$$

where

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

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

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

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

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

Also, by Lemmas 4 and 5,

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

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

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

with probability at least $1 - G^{-2}$. Combining these bounds gives (46) and completes this step.

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

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

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

$$\begin{aligned} (47) \qquad & P\left(\max_{1 \le g \le G} \sqrt{N_g/I_g} \left|\hat{\alpha}_{g,1}(u) - \alpha_{g,1}(u)\right| \le c_{1-\alpha}^M\right) \\ & \le P\left(\max_{1 \le g \le G} \left| \frac{1}{\sqrt{\bar{N}_G}} \sum_{g=1}^{\bar{N}_G} \left(q_{ig} - E_g[q_{ig}]\right) \right| \le c_{1-\alpha}^M + \frac{C \log^{3/4} N_g}{N_g^{1/4}} \right) + o(1). \end{aligned}$$

In turn, since, under our assumptions, $|q_{ig}| \le C$, by Corollary 2.1 in Chernozhukov, Chetverikov, and Kato (2014c), the probability in (47) is bounded from above by

$$
\begin{aligned}
&P\Big(\max_{1\le g\le G} |Y_g| \le c^M_{1-\alpha} + \frac{C\log^{3/4} N_G}{N_G^{1/4}}\Big) + o(1) \\
&\quad \le P\Big(\max_{1\le g\le G} |Y_g| \le c^M_{1-\alpha}\Big) + \frac{C(\log^{3/4} N_G)\cdot(\log^{1/2} G)}{N_G^{1/4}} + o(1) \\
&\quad = 1-\alpha+o(1),
\end{aligned}
$$

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

$$
\begin{aligned}
&P\Big(\max_{1\le g\le G}\sqrt{N_g/I_g}\big|\hat\alpha_{g,1}(u)-\alpha_{g,1}(u)\big| \le c^M_{1-\alpha}\Big) \\
&\quad \le 1-\alpha+o(1).
\end{aligned}
\tag{48}
$$

Similar arguments also give

$$
\begin{aligned}
&P\Big(\max_{1\le g\le G}\sqrt{N_g/I_g}\big|\hat\alpha_{g,1}(u)-\alpha_{g,1}(u)\big| \le c^M_{1-\alpha}\Big) \\
&\quad \ge 1-\alpha-o(1).
\end{aligned}
\tag{49}
$$

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 $c_M$, $c_f$, $C_M$, $C_f$, $C_L$, and $\bar C$. Second, among Lemmas 1–10, Lemmas 4–9 deal with within group variation, and so apply under our conditions without changes. The statement of Lemma 1 holds without changes, but in the proof, Chebyshev's inequality applies on cluster level, that is, for $k = 1, \dots, d_x$ and $l = 1, \dots, d_w$,

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

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

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

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 $G \to \infty$,

$$\frac{1}{G}\sum_{m=1}^{M}\bigg(\sum_{g\in\mathbb{C}_G(m)} \varepsilon_g(u_1)w_g\bigg)\bigg(\sum_{g\in\mathbb{C}_G(m)} \varepsilon_g(u_1)w_g'\bigg) \to_P J^{\mathrm{CS}}(u_1, u_2) \tag{50}$$

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

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

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

$$E\bigg[\bigg|\bigg(\sum_{g\in\mathbb{C}_G(m)} \varepsilon_g(u_1)w_{g,k}\bigg)\bigg(\sum_{g\in\mathbb{C}_G(m)} \varepsilon_g(u_2)w_{g,l}\bigg)\bigg|^{1+\delta}\bigg]$$

$$\leq CE\bigg[\sum_{g,g'\in\mathbb{C}_G(m)} \big|\varepsilon_g(u_1)w_{g,k}\varepsilon_{g'}(u_2)w_{g',l}\big|^{1+\delta}\bigg]$$

$$\leq CE\bigg[\sum_{g,g'\in\mathbb{C}_G(m)} \big(\big|\varepsilon_g(u_1)w_{g,k}\big|^{2+2\delta} + \big|\varepsilon_{g'}(u_2)w_{g',l}\big|^{2+2\delta}\big)\bigg] \leq C,$$

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 $J^{\mathrm{CS}}(u_1, u_2)$. To prove the new statement, first observe that for any finite $\mathcal{U}' \subset \mathcal{U}$,

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

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

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

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

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

and we bound all other terms in the proof similarly. The proof of Theorem 8 is analogous to that of Theorem 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, $N(\varepsilon, T, \rho) = \inf\{\mathrm{Card}(T_\varepsilon) : T_\varepsilon \text{ is an } \varepsilon \text{ net of } T\}$.

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

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

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

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

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

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 $0 < \varepsilon < 1$,*

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

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

PROOF: See Lemma 19.15 in Van der Vaart (1998). *Q.E.D.*

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

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

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

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

*for every $0 < \varepsilon < 1$.*

PROOF: See Lemma A.6 in Chernozhukov, Chetverikov, and Kato (2014a). *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*

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

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

PROOF: See Corollary 5.1 of Chernozhukov, Chetverikov, and Kato (2014a). *Q.E.D.*

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

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

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

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

*for a constant $D$ that depends on $K$ only.*

PROOF: See Theorem 2.14.9 in Van der Vaart and Wellner (1996). *Q.E.D.*

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

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

*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 semi-metric $\rho : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_+$ is called Gaussian if it can be defined as

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

where $G$ is a tight, zero-mean, Gaussian random element in $l^\infty(\mathcal{F})$. 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

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

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 $Z_{n1}, \dots, Z_{nm_n}$ be independent stochastic processes indexed by an arbitrary index set $\mathcal{F}$. Suppose that there exists a Gaussian-dominated semi-metric $\rho$ on $\mathcal{F}$ such that*

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

$$\text{(ii)} \quad \sum_{i=1}^{m_n} E\big[\big(Z_{ni}(f) - Z_{ni}(g)\big)^2\big] \le \rho^2(f, g), \quad \textit{for every } f, g,$$

$$\text{(iii)} \quad \sup_{t>0} \sum_{i=1}^{m_n} t^2 P\Big(\sup_{f,g \in B(\varepsilon)} \big|Z_{ni}(f) - Z_{ni}(g)\big| > t\Big) \le \varepsilon^2,$$

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

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

## ADDITIONAL REFERENCES

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2006): "Conditional Quantile Processes Based on Series or Many Regressors," Working Paper. [16]

CHERNOZHUKOV, V., D. CHETVERIKOV, AND K. KATO (2014a): "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, 42, 1564–1597. [35,36]

——— (2014b): "Anti-Concentration and Honest, Adaptive Confidence Bands," *The Annals of Statistics*, 42, 1787–1818. [28,31]

——— (2014c): "Central Limit Theorems and Bootstrap in High Dimensions," available at arXiv:1412.3661v2. [32]

——— (2015): "Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70. [32]

KOENKER, R. (2005): *Quantile Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press. [6]

LEHMANN, E., AND J. ROMANO (2005): *Testing Statistical Hypotheses*. Springer Texts in Statistics. New York: Springer. [13]

POWELL, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325. [6]

TAO, T. (2012): *Topics in Random Matrix Theory*. Providence, RI: American Mathematical Society. [11]

VAN DER VAART, A. (1998): *Asymptotic Statistics*. Cambridge: Cambridge University Press. [35]

*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.*