Supplementary Appendix to

Wild Bootstrap and Asymptotic Inference with Multiway Clustering

James G. MacKinnon*
Queen's University
jgm@econ.queensu.ca

Morten Ørregaard Nielsen
Queen's University and CREATES
mon@econ.queensu.ca

Matthew D. Webb
Carleton University
matt.webb@carleton.ca

September 13, 2019

A Proofs of Main Results

This supplementary appendix to our paper (MacKinnon, Nielsen, and Webb 2019) contains mathematical proofs of all theorems.

A.1 Proof of Theorem 1

The result in (19) is an immediate consequence of Propositions 4.3 and 4.4 of Davezies et al. (2018), where (16) is assumed. Under (17), there is clustering only at the intersection level, and under (18) there is no clustering. Both of these are special cases of one-way clustering, so that (20) and (21) follow from Djogbenou et al. (2019) after noting that our assumptions imply that the cluster sizes N_{gh} are bounded almost surely.

A.2 Proof of Theorem 2

The results of the theorem follow directly from the definitions of $\hat{\boldsymbol{V}}_2$ and $\hat{\boldsymbol{V}}_3$ in (9) and (6), respectively, and application of Lemma A.1, which is proven in the next subsection. For example, under (16) it follows from this lemma that

$$R(\hat{\boldsymbol{V}}_2 - \hat{\boldsymbol{V}}_3) = R\hat{\boldsymbol{V}}_I = O_P\Big((GH)^{-1}R\Big) \xrightarrow{P} 0.$$

Lemma A.1. *Suppose Assumptions 1–6 are satisfied.*

*Corresponding author. Address: Department of Economics, 94 University Avenue, Queen's University, Kingston, Ontario K7L 3N6, Canada. Email: jgm@econ.queensu.ca. Tel. 613-533-2293. Fax 613-533-6668.

(a) *If (24) holds, so that the DGP is clustered along the first dimension, then*

$$G\hat{\boldsymbol{V}}_G \xrightarrow{P} \boldsymbol{V}_G \quad \textit{and} \quad GH\hat{\boldsymbol{V}}_I \xrightarrow{P} \boldsymbol{V}_I.$$

(b) *If (25) holds, so that the DGP is not clustered along the first dimension, but it is clustered along the second dimension, then*

$$GH\boldsymbol{a}^\top\hat{\boldsymbol{V}}_G\boldsymbol{a} \xrightarrow{d} W_1^2 \quad \textit{and} \quad GH\hat{\boldsymbol{V}}_I \xrightarrow{P} \boldsymbol{V}_I,$$

where W_1^2 is a random variable satisfying $W_1^2 > 0$ almost surely.

(c) *If (17) holds, so that the DGP is clustered by intersections, then*

$$GH\hat{\boldsymbol{V}}_m \xrightarrow{P} \boldsymbol{V}_I \quad \textit{for } m \in \{G, H, I\}.$$

(d) *If (18) holds, so that the DGP is not clustered, then*

$$GH\hat{\boldsymbol{V}}_m \xrightarrow{P} \boldsymbol{V}_I \quad \textit{for } m \in \{G, H, I\}.$$

A.3 Proof of Lemma A.1

Proof for cases (a) and (b): The results in case (a) and the second result in case (b) are given in Proposition 4.4 in Davezies et al. (2018). For the first result in case (b), we use the decomposition $\hat{\boldsymbol{u}}_{gh} = \boldsymbol{u}_{gh} - \boldsymbol{X}_{gh}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ such that $\sum_{h=1}^H \boldsymbol{X}_{gh}^\top\hat{\boldsymbol{u}}_{gh} = \sum_{h=1}^H \boldsymbol{X}_{gh}^\top\boldsymbol{u}_{gh} - (GH)^{-1}\sum_{h=1}^H \boldsymbol{X}_{gh}^\top\boldsymbol{X}_{gh}\boldsymbol{Q}^{-1}\boldsymbol{X}^\top\boldsymbol{u}$, where, under (25), $H^{-1}\sum_{h=1}^H \boldsymbol{X}_{gh}^\top\boldsymbol{X}_{gh} \xrightarrow{P} \boldsymbol{Q}_0$ and $\boldsymbol{Q} \xrightarrow{P} \boldsymbol{Q}_0$; see Assumption 2. Thus, for any arbitrary $\boldsymbol{\eta}$, we write

$$GH\boldsymbol{\eta}^\top\hat{\boldsymbol{\Gamma}}_H\boldsymbol{\eta} = \frac{1}{G}\sum_{g=1}^G\Big(H^{-1/2}\sum_{h=1}^H \boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\hat{\boldsymbol{u}}_{gh}\Big)^2 = \frac{1}{G}\sum_{g=1}^G\Big(H^{-1/2}\sum_{h=1}^H z_{gh}\Big)^2 + o_P(1),$$

where, for any (fixed) g, $z_{gh} = \boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\boldsymbol{u}_{gh} - G^{-1}\boldsymbol{\eta}^\top\boldsymbol{X}_h^\top\boldsymbol{u}_h$ is i.i.d. across h with mean zero and finite variance. For fixed G, it follows that the random vector $H^{-1/2}\sum_{h=1}^H(z_{1h}, \ldots, z_{Gh})^\top$ is asymptotically normal as $H \to \infty$ with mean zero and finite $G \times G$ variance matrix, say $\boldsymbol{J}_G$. Still for fixed G, it follows that $G^{-1}\sum_{g=1}^G(H^{-1/2}\sum_{h=1}^H z_{gh})^2 \xrightarrow{d} G^{-1}\sum_{m=1}^M \nu_m\|\boldsymbol{\mu}_m\|^2 Z_m^2$ as $H \to \infty$, where $(\nu_m, \boldsymbol{\mu}_m)$ denote the eigenvalues and eigenvectors of $\boldsymbol{J}_G$, $M \le G$ is the number of non-zero eigenvalues, and Z_m denote i.i.d. standard normal random variables. Next, $G^{-1}\nu_m\|\mu_m\|^2 \to \omega_m^2 \in [0, \infty)$ for all $m \ge 1$, where $\omega_m > 0$ for at least one m. Hence, $G^{-1}\sum_{g=1}^G(H^{-1/2}\sum_{h=1}^H z_{gh})^2 \xrightarrow{d} \sum_{m=1}^\infty \omega_m^2 Z_m^2$, which is a (scaled) weighted sum of χ_1^2-distributions.

Proof for cases (c) and (d): Under (17), we can apply the results of Djogbenou et al. (2019), for the same reason as in the proof of (20), to conclude that each term in (7), multiplied by GH, converges in probability to $\boldsymbol{\Gamma}_I$ defined in (14). The convergence in probability of $\hat{\boldsymbol{V}}_2$ and $\hat{\boldsymbol{V}}_3$, normalized by GH, follows. Similarly, under (18), each term in (7), multiplied by GH, converges in probability to $\boldsymbol{\Gamma}_I$.

A.4 Proof of Theorem 3

To prove Theorem 3 we first present the bootstrap equivalents of Theorems 1 and 2. These are given in Theorems A.1 and A.2, the proofs of which are in the next subsections.

Theorem A.1. *Suppose Assumptions 1–7 are satisfied and that H_0 is true. Let $m \in \{G, H, I, NC\}$ denote bootstrap clustering by the first dimension, the second dimension, intersections, and individual observations, respectively; c.f. step 3(a). Then it holds that*

$$(\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_m \boldsymbol{a})^{-1/2} \boldsymbol{a}^\top (\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}) \xrightarrow{d^*} Z, \text{ in probability,}$$

where $Z \sim \mathrm{N}(0, 1)$.

Theorem A.2. *Suppose Assumptions 1–7 are satisfied and that H_0 is true.*

(i) Suppose the bootstrap DGP in step 3(a) is clustered along the first (G) dimension (results for bootstrap clustering along the second dimension are symmetric).

(a) If (24) holds, so that the DGP is clustered along the first dimension, then

$$G(\hat{\boldsymbol{V}}_2^* - \ddot{\boldsymbol{V}}_G) \xrightarrow{P^*} \boldsymbol{0} \quad \text{and} \quad G(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_G) \xrightarrow{P^*} \boldsymbol{0}, \text{ in probability.}$$

(b) If (25) holds, so that the DGP is not clustered along the first dimension, but it is clustered along the second dimension, then

$$GH\boldsymbol{a}^\top(\hat{\boldsymbol{V}}_2^* - \ddot{\boldsymbol{V}}_G - \ddot{\boldsymbol{V}}_I)\boldsymbol{a} \xrightarrow{d^*} W_0 \quad \text{and} \quad GH\boldsymbol{a}^\top(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_G)\boldsymbol{a} \xrightarrow{d^*} W_0, \text{ in probability,}$$

where W_0 is a zero mean random variable that is independent of Z in Theorem A.1.

(c) If (17) holds, so that the DGP is clustered by intersections, then

$$GH(\hat{\boldsymbol{V}}_2^* - \ddot{\boldsymbol{V}}_G - \ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0} \quad \text{and} \quad GH(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_G) \xrightarrow{P^*} \boldsymbol{0}, \text{ in probability.}$$

(d) If (18) holds, so that the DGP is not clustered, then

$$GH(\hat{\boldsymbol{V}}_2^* - \ddot{\boldsymbol{V}}_G - \ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0} \quad \text{and} \quad GH(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_G) \xrightarrow{P^*} \boldsymbol{0}, \text{ in probability.}$$

(ii) If the bootstrap DGP in step 3(a) is clustered by intersections, then

$$GH(\hat{\boldsymbol{V}}_2^* - 2\ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0} \quad \text{and} \quad GH(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0}, \text{ in probability.}$$

(iii) If the bootstrap DGP in step 3(a) is the WB, then

$$GH(\hat{\boldsymbol{V}}_2^* - 2\ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0} \quad \text{and} \quad GH(\hat{\boldsymbol{V}}_3^* - \ddot{\boldsymbol{V}}_I) \xrightarrow{P^*} \boldsymbol{0}, \text{ in probability.}$$

Let $m \in \{G, H, I, NC\}$ denote bootstrap clustering by the first dimension, the second dimension, intersections, and individual observations, respectively; c.f. step 3(a). We then decompose the bootstrap t-statistic as

$$t_{a,j}^* = \frac{\boldsymbol{a}^\top(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})}{(\boldsymbol{a}^\top \hat{\boldsymbol{V}}_j^* \boldsymbol{a})^{1/2}} = \left(\frac{\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_m \boldsymbol{a}}{\boldsymbol{a}^\top \hat{\boldsymbol{V}}_j^* \boldsymbol{a}}\right)^{1/2} \frac{\boldsymbol{a}^\top(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})}{(\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_m \boldsymbol{a})^{1/2}} = (A_{m,j}^*)^{1/2} B_m^*, \tag{A.1}$$

say. From Theorem A.1 we find that $B_m^* \xrightarrow{d^*} Z \sim \mathrm{N}(0,1)$, in probability, for all m.

For the first term on the right-hand side of (A.1), the result follows by direct application of Lemma A.1 and Theorem A.2. In particular, for cases (i)(a),(c), and (ii), $A_{m,j}^* \xrightarrow{P^*} q$, in probability, where $q = 1/2$ or $q = 1$ is the variance of the limit distribution of $t_{a,j}^*$. For case (i)(b), we write $A_{m,j}^* = 1 - (\hat{\boldsymbol{V}}_j^* - \ddot{\boldsymbol{V}}_G)/\hat{\boldsymbol{V}}_j^*$ and apply Lemma A.1 and Theorem A.2. Note that, because H_0 is true, the results of Lemma A.1 also apply to the variance estimators imposing the null, i.e. all $\hat{\boldsymbol{V}}$ in Lemma A.1 can be replaced by $\ddot{\boldsymbol{V}}$. The random variable W_1^2 may then be different, but since the explicit form of W_1^2 is not needed, that is not an issue. Finally, Z and W_0 are generated by the bootstrap measure and are both therefore independent of W_1^2.

A.5 Proof of Theorem A.1

We give the proof only for the case where the bootstrap is clustered along the first dimension; that is, $\boldsymbol{u}_g^* = \ddot{\boldsymbol{u}}_g v_g^*$. The proofs for the other cases are entirely analogous. First note that

$$(\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_G \boldsymbol{a})^{-1/2} \boldsymbol{a}^\top (\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}) = \sum_{g=1}^G z_g^*, \quad z_g^* = (\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_G \boldsymbol{a})^{-1/2} \boldsymbol{a}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g v_g^*.$$

Because v_g^* is independent across g with mean zero and variance one, it follows that z_g^* is independent across g with $\mathrm{E}^*(z_g^*) = 0$ and $\mathrm{Var}^*(\sum_{g=1}^G z_g^*) = \sum_{g=1}^G \mathrm{Var}^*(z_g^*) = 1$. The Lyapunov condition is satisfied (with P-probability converging to one) because

$$\begin{aligned}\sum_{g=1}^G \mathrm{E}^*|z_g^*|^4 &\le \mathrm{E}(v^{*4}) (\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_G \boldsymbol{a})^{-2} \|\boldsymbol{Q}^{-1}\|^4 \sum_{g=1}^G \Big\| \frac{1}{GH} \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g \Big\|^4 \\ &= O_P(1) (\boldsymbol{a}^\top \ddot{\boldsymbol{V}}_G \boldsymbol{a})^{-2} \frac{1}{(GH)^4} \sum_{g=1}^G \Big\| \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g \Big\|^4 \xrightarrow{P} 0,\end{aligned}$$

regardless of the clustering structure in the DGP. To see this, suppose first that (24) holds, in which case the DGP is clustered along the first (G) dimension. Then $\boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g = \sum_{h=1}^H \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh}$ is of order $O_P(H)$ and $G\ddot{\boldsymbol{V}}_G \xrightarrow{P} \boldsymbol{V}_G > 0$; see Davezies et al. (2018) and Lemma A.1. However, if the DGP is not clustered along the first dimension (under (17), (18), or (25)), then $\boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g = O_P(H^{1/2})$ and $\ddot{\boldsymbol{V}}_G^{-1} = O_P(GH)$; see also Djogbenou et al. (2019) and Lemma A.1. In either case, $\sum_{g=1}^G \mathrm{E}^*|z_g^*|^4 = O_P(G^{-1})$.

A.6 Proof of Theorem A.2

In all cases, the factors $\boldsymbol{Q}^{-1}$ in the definitions of $\hat{\boldsymbol{V}}_j^*$ are functions only of the original data and satisfy $\boldsymbol{Q} \xrightarrow{P} \boldsymbol{Q}_0 > 0$. Hence, these factors have no impact on the proofs. We therefore prove most results for the corresponding $\hat{\boldsymbol{\Gamma}}_m^*$; see (7) and (22). Specifically, we prove the following lemma, which suffices for the theorem.

Lemma A.2. *Suppose Assumptions 1–7 are satisfied and that H_0 is true.*

(i) Suppose the bootstrap DGP in step 3(a) is clustered along the first (G) dimension (results for bootstrap clustering along the second dimension are symmetric).

(a) If (24) holds, so that the DGP is clustered along the first dimension, then

$$G(\hat{\mathbf{\Gamma}}^*_G - \ddot{\mathbf{\Gamma}}_G) \xrightarrow{P^*} \mathbf{0} \quad \textit{and} \quad GH(\hat{\mathbf{\Gamma}}^*_I - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability},$$
$$\mathrm{Var}^*(GH(\hat{\mathbf{\Gamma}}^*_H - \ddot{\mathbf{\Gamma}}_I)) = O_P(1).$$

(b) If (25) holds, so that the DGP is not clustered along the first dimension, but it is clustered along the second dimension, then

$$GH(\hat{\mathbf{\Gamma}}^*_G - \ddot{\mathbf{\Gamma}}_G) \xrightarrow{P^*} \mathbf{0} \quad \textit{and} \quad GH(\hat{\mathbf{\Gamma}}^*_I - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability},$$
$$GH\boldsymbol{a}^\top(\hat{\boldsymbol{V}}^*_H - \ddot{\boldsymbol{V}}_I)\boldsymbol{a} \xrightarrow{d^*} W_0, \ \textit{in probability},$$

where W_0 is a zero mean random variable that is independent of Z in Theorem A.1.

(c) If (17) holds, so that the DGP is clustered by intersections, then

$$GH(\hat{\mathbf{\Gamma}}^*_G - \ddot{\mathbf{\Gamma}}_G) \xrightarrow{P^*} \mathbf{0} \quad \textit{and} \quad GH(\hat{\mathbf{\Gamma}}^*_m - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability, for } m \in \{H, I\}.$$

(d) If (18) holds, so that the DGP is not clustered, then

$$GH(\hat{\mathbf{\Gamma}}^*_G - \ddot{\mathbf{\Gamma}}_G) \xrightarrow{P^*} \mathbf{0} \quad \textit{and} \quad GH(\hat{\mathbf{\Gamma}}^*_m - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability, for } m \in \{H, I\}.$$

(ii) If the bootstrap DGP in step 3(a) is clustered by intersections, then

$$GH(\hat{\mathbf{\Gamma}}^*_m - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability, for } m \in \{G, H, I\}.$$

(iii) If the bootstrap DGP in step 3(a) is the WB, then

$$GH(\hat{\mathbf{\Gamma}}^*_m - \ddot{\mathbf{\Gamma}}_I) \xrightarrow{P^*} \mathbf{0}, \ \textit{in probability, for } m \in \{G, H, I\}.$$

A.7 Proof of Lemma A.2

We prove convergence in mean square. That is, we show that the second moment (conditional on the sample) converges to zero (in P-probability). Let $\boldsymbol{\eta}$ be an arbitrary conforming vector.

Proof for case (i): First, using the decomposition $\hat{\boldsymbol{u}}^*_g = \ddot{\boldsymbol{u}}_g v^*_g - \boldsymbol{X}_g(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})$, we find that

$$\boldsymbol{\eta}^\top(\hat{\mathbf{\Gamma}}^*_G - \ddot{\mathbf{\Gamma}}_G)\boldsymbol{\eta} = \frac{1}{(GH)^2}\sum_{g=1}^{G} \boldsymbol{\eta}^\top \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g \ddot{\boldsymbol{u}}_g^\top \boldsymbol{X}_g \boldsymbol{\eta}(v^{*2}_g - 1) \tag{A.2}$$
$$- \frac{2}{(GH)^2}\sum_{g=1}^{G} \boldsymbol{\eta}^\top \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g (\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})^\top \boldsymbol{X}_g^\top \boldsymbol{X}_g \boldsymbol{\eta} v^*_g \tag{A.3}$$
$$+ \frac{1}{(GH)^2}\sum_{g=1}^{G} \Big(\boldsymbol{\eta}^\top \boldsymbol{X}_g^\top \boldsymbol{X}_g (\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})\Big)^2. \tag{A.4}$$

Because $(v^{*2}_g - 1)$ is independent and identically distributed across g with mean zero and finite variance, the conditional second moment of (A.2) is

$$\mathrm{E}^*((\text{A.2})^2) = \frac{1}{(GH)^4}\mathrm{E}^*\Big((v^{*2} - 1)^2\Big)\sum_{g=1}^{G}(\boldsymbol{\eta}^\top \boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g \ddot{\boldsymbol{u}}_g^\top \boldsymbol{X}_g \boldsymbol{\eta})^2.$$

Under (24), where the DGP is clustered along the first (G) dimension, $\boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g = \sum_{h=1}^H \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh}$ is of order $O_P(H)$. However, if the DGP is not clustered along the first dimension (under (17), (18), or (25)), then $\boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g = O_P(H^{1/2})$. This shows the results for (A.2) for case (i). The conditional second moment of (A.3) is

$$\mathrm{E}^*((\text{A.3})^2) = \frac{4}{(GH)^4}\mathrm{E}^*\Big(\sum_{g_1,g_2=1}^G \boldsymbol{\eta}^\top \boldsymbol{X}_{g_1}^\top \ddot{\boldsymbol{u}}_{g_1} \ddot{\boldsymbol{u}}_{g_2}^\top \boldsymbol{X}_{g_2} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}_{g_1}^\top \boldsymbol{X}_{g_1} \boldsymbol{\eta} v_{g_1}^* v_{g_2}^*\Big)^2,$$

where we note that expanding the square results in four summations, but two of these are eliminated because v_g^* is independent across g, so that the summation indexes must be equal in pairs. Using this together with the aforementioned orders of magnitude of $\boldsymbol{X}_g^\top \ddot{\boldsymbol{u}}_g$ and the facts that $(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = O_P((GH)^{-1})$ and $\boldsymbol{X}_g^\top \boldsymbol{X}_g = O_P(H)$ (Davezies et al. 2018) yields the desired results for (A.3) for case (i). Finally, (A.4) is a non-negative random variable, and noting that $\mathrm{Var}^*(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}}) = \ddot{\boldsymbol{V}}_G$, its conditional mean is

$$\mathrm{E}^*((\text{A.4})) = \frac{1}{(GH)^2}\sum_{g=1}^G \boldsymbol{\eta}^\top \boldsymbol{X}_g^\top \boldsymbol{X}_g \ddot{\boldsymbol{V}}_G \boldsymbol{X}_g^\top \boldsymbol{X}_g \boldsymbol{\eta} = O_P(G^{-1}\|\ddot{\boldsymbol{V}}_G\|).$$

The results for (A.4) for case (i) then follow by application of Lemma A.1.

Next, we find that

$$GH\boldsymbol{\eta}^\top(\hat{\boldsymbol{\Gamma}}_I^* - \ddot{\boldsymbol{\Gamma}}_I)\boldsymbol{\eta} = \frac{1}{GH}\sum_{g=1}^G\sum_{h=1}^H \boldsymbol{\eta}^\top \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh} \ddot{\boldsymbol{u}}_{gh}^\top \boldsymbol{X}_{gh} \boldsymbol{\eta}(v_g^{*2} - 1) \tag{A.5}$$

$$- \frac{2}{GH}\sum_{g=1}^G\sum_{h=1}^H \boldsymbol{\eta}^\top \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh}(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})^\top \boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh} \boldsymbol{\eta} v_g^* \tag{A.6}$$

$$+ \frac{1}{GH}\sum_{g=1}^G\sum_{h=1}^H \Big(\boldsymbol{\eta}^\top \boldsymbol{X}_{gh}^\top \boldsymbol{X}_{gh}(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})\Big)^2. \tag{A.7}$$

The proofs for each of the terms (A.5)–(A.7) are nearly identical to those for (A.2)–(A.4), and they are therefore omitted.

For $\hat{\boldsymbol{\Gamma}}_H^*$ we find that

$$GH\boldsymbol{\eta}^\top \hat{\boldsymbol{\Gamma}}_H^* \boldsymbol{\eta} = \frac{1}{GH}\sum_{h=1}^H (\boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \hat{\boldsymbol{u}}_h^*)^2 = \frac{1}{GH}\sum_{h=1}^H \Big(\boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \boldsymbol{u}_h^* - \boldsymbol{X}_h(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})\Big)^2$$

$$= \frac{1}{GH}\sum_{h=1}^H (\boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \boldsymbol{u}_h^*)^2 + A^*, \tag{A.8}$$

where $\mathrm{E}^*|A^*| = O_P(G\|\ddot{\boldsymbol{V}}_G\|)$, in probability, by application of Lemma A.1 and the Cauchy-Schwarz inequality, because

$$\mathrm{E}^*\Big(\frac{1}{GH}\sum_{h=1}^H (\boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \boldsymbol{X}_h(\hat{\boldsymbol{\beta}}^* - \ddot{\boldsymbol{\beta}})\Big)^2 = \frac{1}{GH}\sum_{h=1}^H \boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \boldsymbol{X}_h \ddot{\boldsymbol{V}}_G \boldsymbol{X}_h^\top \boldsymbol{X}_h \boldsymbol{\eta} = O_P(G\|\ddot{\boldsymbol{V}}_G\|).$$

This shows that A^* is of the required order of magnitude in part (a) and is negligible in parts (b)–(d). Noting that $\boldsymbol{X}_h^\top \boldsymbol{u}_h^* = \sum_{g=1}^G \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh} v_g^*$, the main term in (A.8) satisfies

$$\frac{1}{GH}\sum_{h=1}^H (\boldsymbol{\eta}^\top \boldsymbol{X}_h^\top \boldsymbol{u}_h^*)^2 - GH\boldsymbol{\eta}^\top \ddot{\boldsymbol{\Gamma}}_I \boldsymbol{\eta} = \frac{1}{GH}\sum_{h=1}^H\sum_{g=1}^G \boldsymbol{\eta}^\top \boldsymbol{X}_{gh}^\top \ddot{\boldsymbol{u}}_{gh} \ddot{\boldsymbol{u}}_{gh}^\top \boldsymbol{X}_{gh} \boldsymbol{\eta}(v_g^{*2} - 1) \tag{A.9}$$

$$+\frac{1}{GH}\sum_{h=1}^{H}\sum_{g_1\neq g_2}^{G}\boldsymbol{\eta}^\top\boldsymbol{X}_{g_1h}^\top\ddot{\boldsymbol{u}}_{g_1h}\ddot{\boldsymbol{u}}_{g_2h}^\top\boldsymbol{X}_{g_2h}\boldsymbol{\eta}v_{g_1}^*v_{g_2}^*. \quad \text{(A.10)}$$

By independence of $(v_g^{*2}-1)$ across g, it is easily seen that $\mathrm{E}^*((\text{A.9})^2) = O_P(G^{-1})$, showing the results for (A.9) for case (i). Similarly,

$$\mathrm{E}^*((\text{A.10})^2) = \frac{2}{(GH)^2}\sum_{g_1\neq g_2}^{G}\Big(\sum_{h=1}^{H}\boldsymbol{\eta}^\top\boldsymbol{X}_{g_1h}^\top\ddot{\boldsymbol{u}}_{g_1h}\ddot{\boldsymbol{u}}_{g_2h}^\top\boldsymbol{X}_{g_2h}\boldsymbol{\eta}\Big)^2,$$

which is $O_P(1)$ in part (a) and $o_P(1)$ in parts (c) and (d), showing the results for (A.10) for those parts. Thus, only part (b) remains for (A.10). For any fixed h, as $G\to\infty$,

$$\frac{1}{G^{1/2}}\boldsymbol{\eta}^\top\boldsymbol{X}_h^\top\boldsymbol{u}_h^* = \frac{1}{G^{1/2}}\sum_{g=1}^{G}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\ddot{\boldsymbol{u}}_{gh}v_g^* \xrightarrow{d^*} \mathrm{N}\Big(0, \operatorname*{plim}_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\ddot{\boldsymbol{u}}_{gh})^2\Big), \quad \text{(A.11)}$$

in probability. Moreover, for fixed $h_1\neq h_2$, as $G\to\infty$,

$$\mathrm{E}^*\Big(\frac{1}{G}\boldsymbol{\eta}^\top\boldsymbol{X}_{h_1}^\top\boldsymbol{u}_{h_1}^*\boldsymbol{\eta}^\top\boldsymbol{X}_{h_2}^\top\boldsymbol{u}_{h_2}^*\Big) = \frac{1}{G}\sum_{g=1}^{G}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh_1}^\top\ddot{\boldsymbol{u}}_{gh_1}\ddot{\boldsymbol{u}}_{gh_2}^\top\boldsymbol{X}_{gh_2}\boldsymbol{\eta} \xrightarrow{P} 0. \quad \text{(A.12)}$$

It follows from (A.11), (A.12), and the continuous mapping theorem that, for fixed H, as $G\to\infty$,

$$\frac{1}{GH}\sum_{h=1}^{H}(\boldsymbol{\eta}^\top\boldsymbol{X}_h^\top\boldsymbol{u}_h^*)^2 \xrightarrow{d^*} \frac{1}{H}\sum_{h=1}^{H}\Big(\operatorname*{plim}_{G\to\infty}\frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\ddot{\boldsymbol{u}}_{gh})^2\Big)Z_h^2, \text{ in probability,}$$

where $Z_h \sim i.i.d.\mathrm{N}(0,1)$ for $h=1,\ldots,H$. Because $(GH)^{-1}\sum_{g=1}^{G}\sum_{h=1}^{H}(\boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\ddot{\boldsymbol{u}}_{gh})^2 \xrightarrow{P} \boldsymbol{\eta}^\top\boldsymbol{\Gamma}_I\boldsymbol{\eta}<\infty$ by Lemma A.1, it follows that

$$\frac{1}{GH}\sum_{h=1}^{H}(\boldsymbol{\eta}^\top\boldsymbol{X}_h^\top\boldsymbol{u}_h^*)^2 \xrightarrow{d^*} \sum_{m=1}^{\infty}\upsilon_m^2Z_m^2, \text{ in probability,} \quad \text{(A.13)}$$

where $Z_m \sim i.i.d.\mathrm{N}(0,1)$ for $m=1,2,\ldots$. The right-hand side of (A.13) is a weighted sum of χ_1^2-distributions, where the weights satisfy $\sum_{m=1}^{\infty}\upsilon_m^2 = \boldsymbol{\eta}^\top\boldsymbol{\Gamma}_I\boldsymbol{\eta}$. Hence, using $\boldsymbol{Q}\xrightarrow{P}\boldsymbol{Q}_0$ and combining (A.8), (A.9), (A.10), and (A.13), we find for part (b) that

$$GH\boldsymbol{a}^\top(\hat{\boldsymbol{V}}_H^* - \ddot{\boldsymbol{V}}_I)\boldsymbol{a} \xrightarrow{d^*} \sum_{m=1}^{\infty}\tau_m^2(Z_m^2-1) = W_0, \text{ in probability,}$$

where the weights τ_m are derived from υ_m by setting $\boldsymbol{\eta}=\boldsymbol{Q}_0^{-1}\boldsymbol{a}$ in the latter and the τ_m thus satisfy $\sum_{m=1}^{\infty}\tau_m^2 = \boldsymbol{a}^\top\boldsymbol{V}_I\boldsymbol{a}$. Finally, W_0 is independent of Z because

$$\mathrm{E}^*\Big((\boldsymbol{a}^\top\ddot{\boldsymbol{V}}_G\boldsymbol{a})^{-1/2}\boldsymbol{a}^\top\boldsymbol{Q}^{-1}\frac{1}{GH}\sum_{g_1=1}^{G}\boldsymbol{X}_{g_1}^\top\ddot{\boldsymbol{u}}_{g_1}v_{g_1}^*\frac{1}{G}\Big(\sum_{g_2=1}^{G}\boldsymbol{\eta}^\top\boldsymbol{X}_{g_2h}^\top\ddot{\boldsymbol{u}}_{g_2h}v_{g_2}^*\Big)^2\Big) = O_P((GH)^{-1/2})$$

using Lemma A.1, independence of v_g^* across g (to eliminate the summation over g_2), and the fact that $\boldsymbol{X}_g^\top\ddot{\boldsymbol{u}}_g = O_P(H^{1/2})$ under (25).

Proof for case (ii): First, we find that

$$GH\boldsymbol{\eta}^\top(\hat{\boldsymbol{\Gamma}}_G^* - \ddot{\boldsymbol{\Gamma}}_I)\boldsymbol{\eta} = \frac{1}{GH}\sum_{g=1}^{G}\sum_{h=1}^{H}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh}^\top\ddot{\boldsymbol{u}}_{gh}\ddot{\boldsymbol{u}}_{gh}^\top\boldsymbol{X}_{gh}\boldsymbol{\eta}(v_{gh}^{*2}-1) \tag{A.14}$$

$$+\frac{1}{GH}\sum_{g=1}^{G}\sum_{h_1\neq h_2}^{H}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh_1}^\top\ddot{\boldsymbol{u}}_{gh_1}\ddot{\boldsymbol{u}}_{gh_2}^\top\boldsymbol{X}_{gh_2}\boldsymbol{\eta}v_{gh_1}^*v_{gh_2}^* \tag{A.15}$$

$$-\frac{2}{GH}\sum_{g=1}^{G}\sum_{h_1,h_2=1}^{H}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh_1}^\top\ddot{\boldsymbol{u}}_{gh_1}(\hat{\boldsymbol{\beta}}^*-\ddot{\boldsymbol{\beta}})^\top\boldsymbol{X}_{gh_2}^\top\boldsymbol{X}_{gh_2}\boldsymbol{\eta}v_{gh_1}^*v_{gh_2}^* \tag{A.16}$$

$$+\frac{1}{GH}\sum_{g=1}^{G}\sum_{h_1,h_2=1}^{H}\boldsymbol{\eta}^\top\boldsymbol{X}_{gh_1}^\top\boldsymbol{X}_{gh_1}(\hat{\boldsymbol{\beta}}^*-\ddot{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}^*-\ddot{\boldsymbol{\beta}})^\top\boldsymbol{X}_{gh_2}^\top\boldsymbol{X}_{gh_2}\boldsymbol{\eta}v_{gh_1}^*v_{gh_2}^*. \tag{A.17}$$

The proofs for (A.14), (A.16), and (A.17) are nearly identical to those for (A.2)–(A.4), and are therefore omitted. For (A.15) we find

$$\mathrm{E}^*((\text{A.15})^2) = \frac{1}{(GH)^2}\sum_{g_1,g_2}^{G}\sum_{h_1\neq h_1'}^{H}\sum_{h_2\neq h_2'}^{H}t(g_1,g_2,h_1,h_1',h_2,h_2')\mathrm{E}^*(v_{g_1h_1}^*v_{g_1h_1'}^*v_{g_2h_2}^*v_{g_2h_2'}^*), \tag{A.18}$$

where $t(g_1,g_2,h_1,h_1',h_2,h_2') = \boldsymbol{\eta}^\top\boldsymbol{X}_{g_1h_1}^\top\ddot{\boldsymbol{u}}_{g_1h_1}\ddot{\boldsymbol{u}}_{g_1h_1'}^\top\boldsymbol{X}_{g_1h_1'}\boldsymbol{\eta}\boldsymbol{\eta}^\top\boldsymbol{X}_{g_2h_2}^\top\ddot{\boldsymbol{u}}_{g_2h_2}\ddot{\boldsymbol{u}}_{g_2h_2'}^\top\boldsymbol{X}_{g_2h_2'}\boldsymbol{\eta}$ is a function only of the original data and is $O_P(1)$. By independence of v_{gh}^* across both g and h, the right-hand side of (A.18) is non-zero only if $g_1 = g_2$ and either $h_1 = h_2, h_1' = h_2'$ or $h_1 = h_2', h_1' = h_2$. In either situation, one summation over g and two summations over h are eliminated, so that (A.18) is at most $O_P(G^{-1})$, which proves the result for $\hat{\boldsymbol{\Gamma}}_G^*$.

The proof for $\hat{\boldsymbol{\Gamma}}_H^*$ is identical to that for $\hat{\boldsymbol{\Gamma}}_G^*$ after interchanging the g and h subscripts throughout. Finally, $GH\boldsymbol{\eta}^\top(\hat{\boldsymbol{\Gamma}}_I^* - \ddot{\boldsymbol{\Gamma}}_I)\boldsymbol{\eta}$ is equal to the sum of (A.14), (A.16), and (A.17), with $h_1 = h_2$ in the latter two, so we have already proven the required result for this term.

Proof for case (iii): The proofs for case (iii) are nearly identical to those for case (ii) and are therefore omitted.

References

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2018). Asymptotic results under multiway clustering. ArXiv e-prints, CREST.

Djogbenou, A. A., J. G. MacKinnon, and M. Ø. Nielsen (2019). Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics 212*, 393–412.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb (2019). Wild bootstrap and asymptotic inference with multiway clustering. *Journal of Business and Economic Statistics*, to appear.