Mathematical Appendix for: Semiparametric Estimation of a Simultaneous Game with Incomplete Information

Andres Aradillas-Lopez*

*The results presented here benefited greatly from discussions with Professor James L. Powell, to whom I am greatly indebted. All remaining errors are tesponsibility of the author. email: aradilla@econ.berkeley.edu. Correspondence: Andres Aradillas-Lopez, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720-3880

A Mathematical Proofs

A.1 Lemmas 3.2, 3.3

Proof of Lemma 3.2: This is a direct consequence of Brouwer's Fixed Point Theorem:[20] Take any $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and any $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$. If $(\widetilde{S1})$-$(\widetilde{S2})$ are satisfied, then for all $(\pi_1, \pi_2) \in \mathbb{R}^2$ we get that $\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ are continuous and strictly bounded inside $[0,1]^2$, which is a compact, convex, nonempty subset of $\mathbb{R}^2$. Therefore, if we restrict the domain to $(\pi_1, \pi_2) \in [0,1]^2$ then all the conditions of Brouwer's Fixed Point Theorem are satisfied and the system must have a fixed point in $[0,1] \times [0,1]$. In addition, because both $\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$ are strictly inside $[0,1]$ for all $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$, $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$ and $(\pi_1, \pi_2) \in \mathbb{R}^2$ then all fixed points must be strictly inside $[0,1]^2$ for all $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$. $\square$

Proof of Lemma 3.3: Fix $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$ and $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$. Then, for any $(\pi_1, \pi_2) \in \mathbb{R}^2$ define:

$$\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) = E\big[G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big]; \qquad \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) = E\big[G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big]$$
$$\delta_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) = E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big]; \qquad \delta_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) = E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big]$$

We will now analyze the cases $\alpha_1 \times \alpha_2 = 0$ and $\alpha_1 \times \alpha_2 \neq 0$ separately.

Case 1: $\boldsymbol{\alpha_1 \times \alpha_2 = 0}$

Suppose $\alpha_1 = 0$ and define $\pi_1^* \equiv E\big[G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1) \mid \boldsymbol{Z}\big]$. Then we trivially have $\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) = \pi_1^*$ for all $\pi_2 \in \mathbb{R}$. Now let $\pi_2^* = \varphi_2(\pi_1^* \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$. Then (π_1^*, π_2^*) is the unique solution to the equilibrium system (1). If $\alpha_2 = 0$ but $\alpha_1 \neq 0$ then the unique equilibrium (π_1^*, π_2^*) would be given by: $\pi_2^* \equiv E\big[G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2) \mid \boldsymbol{Z}\big]$ and $\pi_1^* = \varphi_1(\pi_2^* \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$. These two cases together show that if $\alpha_1 \times \alpha_2 = 0$, then the solution to (1) is unique.

Case 2: $\boldsymbol{\alpha_1 \times \alpha_2 \neq 0}$

If assumptions $(\widetilde{S1})$-$(\widetilde{S2})$ are satisfied and $\alpha_1 \times \alpha_2 \neq 0$, then $\varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$ and $\varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$

[20]See Theorem M.I.1 in Mas-Collel, Whinston and Green (1995).

are continuous, monotonic, one-to-one functions of π_2 and π_1 respectively. Now define the inverse function φ_1^{-1} that satisfies:

$$\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) = \pi_2 \quad \text{if and only if} \quad \varphi_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) = \pi_1$$

Then φ_1^{-1} is well defined and continuous for all $\pi_1 \in (0,1)$. In addition, π_1^* is a solution (for π_1) to the equilibrium system (1) if and only if $\varphi_2(\pi_1^* \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) = \varphi_1^{-1}(\pi_1^* \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$. To show uniqueness of equilibrium, all we need to do is show that π_1^* is unique: In equilibrium, π_2^* must satisfy $\pi_2^* = \varphi_2(\pi_1^* \mid \boldsymbol{Z}, \boldsymbol{\theta}_2)$; since φ_2 is a one-to-one function, then π_1^* implies uniqueness of π_2^*. Therefore, we will focus on π_1^* and define:

$$\Gamma(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}) = \varphi_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) - \varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)$$

then π_1^* is a solution (for π_1) to the equilibrium system (1) if and only if $\Gamma(\pi_1^* \mid \boldsymbol{Z}, \boldsymbol{\theta}) = 0$. Using the properties of inverse functions, we have:

$$\frac{d\Gamma(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta})}{d\pi_1} = \alpha_2 \delta_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) - \frac{1}{\alpha_1 \delta_1(\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)}$$

We will divide the case $\alpha_1 \times \alpha_2 \neq 0$ into two cases: $\alpha_1 \times \alpha_2 < 0$ and $\alpha_1 \times \alpha_2 > 0$ and analyze each one separately.

Case 2.1: $\boldsymbol{\alpha_1 \times \alpha_2 < 0}$

Before proceeding, note that if assumptions $(\widetilde{S1})$-$(\widetilde{S2})$ are satisfied, then $0 < \delta_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1) < \overline{g}_1$ and $0 < \delta_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) < \overline{g}_2$ for all $(\pi_1, \pi_2) \in \mathbb{R}^2$, all $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and all $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$. Therefore, we have:

$$\text{If } \alpha_1 > 0,\ \alpha_2 < 0 \text{ then:} \quad \frac{d\Gamma(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta})}{d\pi_1} < 0 \quad \text{for all } \pi_1 \in \mathbb{R}$$
$$\text{If } \alpha_1 < 0,\ \alpha_2 > 0 \text{ then:} \quad \frac{d\Gamma(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta})}{d\pi_1} > 0 \quad \text{for all } \pi_1 \in \mathbb{R}$$

Therefore, if $\alpha_1 \times \alpha_2 < 0$ then:

$$\text{Sign}\left(\frac{d\Gamma(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta})}{d\pi_1}\right) \quad \text{is constant and different from zero for all } \pi_1 \in \mathbb{R}$$

Thus, if $\alpha_1 \times \alpha_2 < 0$ then $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})$ is a monotonic function of π_1 for all $\pi_1 \in \mathbb{R}$, which means that there is at most one π_1^* such that $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}) = 0$. From the proof of Lemma 3.2 we know that there must exist at least one such π_1^*. This shows that if $\alpha_1 \times \alpha_2 < 0$ then there is a unique π_1^* for which $\Gamma(\pi_1^* \mid \boldsymbol{Z},\boldsymbol{\theta}) = 0$.

Case 2.2: $\boldsymbol{\alpha_1 \times \alpha_2 > 0}$

Define: $\pi_1^{(0)} \equiv \varphi_1(0 \mid \boldsymbol{Z},\boldsymbol{\theta}_1)$ and $\pi_1^{(1)} \equiv \varphi_1(1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1)$. Then, since both $\varphi_1(\pi_2 \mid \boldsymbol{Z},\boldsymbol{\theta}_1)$ and $\varphi_2(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1)$ are strictly inside $[0,1]$ for all $(\pi_1,\pi_2) \in \mathbb{R}^2$, we have that $\pi_1^{(0)} \in (0,1)$, $\pi_1^{(1)} \in (0,1)$ and:

$$\varphi_2(\pi_1^{(0)} \mid \boldsymbol{Z},\boldsymbol{\theta}_2) > \varphi_1^{-1}(\pi_1^{(0)} \mid \boldsymbol{Z},\boldsymbol{\theta}_1) = 0 \quad \text{and} \quad \varphi_2(\pi_1^{(1)} \mid \boldsymbol{Z},\boldsymbol{\theta}_2) < \varphi_1^{-1}(\pi_1^{(1)} \mid \boldsymbol{Z},\boldsymbol{\theta}_1) = 1$$

and therefore $\Gamma(\pi_1^{(0)} \mid \boldsymbol{Z},\boldsymbol{\theta}) > 0$ and $\Gamma(\pi_1^{(1)} \mid \boldsymbol{Z},\boldsymbol{\theta}) < 0$. Now, note that all equilibrium solutions π_1^* must be strictly between $\pi_1^{(0)}$ and $\pi_1^{(1)}$: If $\alpha_1 > 0$ then $\pi_1^{(0)} < \pi_1^{(1)}$ and $\pi_1^* \in (\pi_1^{(0)},\pi_1^{(1)})$, and if $\alpha_1 < 0$ then $\pi_1^{(0)} > \pi_1^{(1)}$ and $\pi_1^* \in (\pi_1^{(1)},\pi_1^{(0)})$. To see why, note that if $\alpha_1 > 0$ then $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) < 0$ for all $\pi_1 \in (0,\pi_1^{(0)})$ and $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) > 1$ for all $\pi_1 \in (\pi_1^{(1)},1)$, whereas if $\alpha_1 < 0$ then $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) < 0$ for all $\pi_1 \in (\pi_1^{(0)},1)$ and $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) > 1$ for all $\pi_1 \in (0,\pi_1^{(1)})$. All these cases are incompatible with an equilibrium since in all of them we have either $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}) > 0$ or $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}) < 0$. Therefore, to prove uniqueness of equilibrium, it is sufficient to show that $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})$ is a monotonic function of π_1 everywhere between $\pi_1^{(0)}$ and $\pi_1^{(1)}$. Suppose that:

$$\alpha_1\alpha_2\delta_1(\pi_2 \mid \boldsymbol{Z},\boldsymbol{\theta}_1)\delta_2(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_2) < 1 \quad \text{for all } (\pi_1,\pi_2) \in [0,1]^2 \qquad (\star)$$

Then, since $\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) \in (0,1)$ for all π_1 between $\pi_1^{(0)}$ and $\pi_1^{(1)}$, we get that $(\star)$ implies that:

$$\alpha_1\alpha_2\delta_1\big(\varphi_1^{-1}(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) \mid \boldsymbol{Z},\boldsymbol{\theta}_1\big)\delta_2\big(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_2\big) < 1 \quad \text{for all } \pi_1 \text{ between } \pi_1^{(0)} \text{ and } \pi_1^{(1)}$$

and therefore:

$$\text{If } \alpha_1 > 0,\, \alpha_2 > 0 \text{ then:} \qquad \frac{d\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})}{d\pi_1} < 0 \quad \text{for all } \pi_1 \in (\pi_1^{(0)}, \pi_1^{(1)})$$

$$\text{If } \alpha_1 < 0,\, \alpha_2 < 0 \text{ then:} \qquad \frac{d\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})}{d\pi_1} > 0 \quad \text{for all } \pi_1 \in (\pi_1^{(1)}, \pi_1^{(0)})$$

Then, if $\alpha_1 \times \alpha_2 > 0$, $(\widetilde{S1})$-$(\widetilde{S2})$ and $(\star)$ hold, then:

$$\text{Sign}\left(\frac{d\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})}{d\pi_1}\right) \quad \text{is constant and different from zero for all } \pi_1 \text{ between } \pi_1^{(0)} \text{ and } \pi_1^{(1)}$$

and therefore $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta})$ is monotonic everywhere between $\pi_1^{(0)}$ and $\pi_1^{(1)}$. Since all equilibria must lie strictly inside this interval, this means that there is at most one π_1^* such that $\Gamma(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}) = 0$. From the proof of Lemma 3.2 we know that there must exist at least one such π_1^*. This shows that if $\alpha_1 \times \alpha_2 > 0$ and $(\widetilde{S1})$-$(\widetilde{S2})$ along with $(\star)$ hold, then there is a unique π_1^* for which $\Gamma(\pi_1^* \mid \boldsymbol{Z},\boldsymbol{\theta}) = 0$.

To complete the proof, we only have to put together cases 1 and 2: Note that if $\alpha_1 \times \alpha_2 \leq 0$ then $(\star)$ holds trivially. In fact, in this case we showed uniqueness of equilibrium without having to use $(\star)$. To show uniqueness we only needed to impose $(\star)$ for the case $\alpha_1 \times \alpha_2 > 0$. Therefore, we can conveniently summarize these results as: "Take $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and suppose assumptions $(\widetilde{S1})$ and $(\widetilde{S2})$ are satisfied. In addition, suppose:

$$\alpha_1\alpha_2 E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z}\big] E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z}\big] < 1 \;\forall (\pi_1,\pi_2) \in [0,1]^2$$

then the equilibrium $\big(\pi_1^*(\boldsymbol{Z},\boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z},\boldsymbol{\theta})\big)$ is unique." This proves the first part of the statement in Lemma 3.3. Uniqueness of equilibrium yields existence of $\mathcal{F}(y_1, y_2 \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$. To show that the latter is a continuous function of the parameters around a neighborhood of this $\boldsymbol{Z}$ and for all $\boldsymbol{X}$ we just have to show that $\big(\pi_1^*, \pi_2^*\big)$ is a continuous function of $\boldsymbol{\theta}$ inside a neighborhood of $\boldsymbol{Z}$. We use $(\star)$ along with the Implicit Function Theorem (IFT) to show this: To show that the equilibrium $\big(\pi_1^*(\boldsymbol{Z},\boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z},\boldsymbol{\theta})\big)$ is a $\mathcal{C}^1$ function, note that the Jacobian of the equilibrium system (1) with respect to π_1 and π_2 is given by:

$$J = \begin{pmatrix} 1 & -\alpha_1\delta_1(\pi_2 \mid \boldsymbol{Z},\boldsymbol{\theta}_1) \\ -\alpha_2\delta_2(\pi_1 \mid \boldsymbol{Z},\boldsymbol{\theta}_2) & 1 \end{pmatrix}$$

which has full-rank if and only if $1 - \alpha_1\alpha_2\delta_1(\pi_2 \mid \boldsymbol{Z}, \boldsymbol{\theta}_1)\delta_2(\pi_1 \mid \boldsymbol{Z}, \boldsymbol{\theta}_2) \neq 0$. Therefore, if the assumption of Lemma 3.3 (i.e ($\star$)) is satisfied, then J has full-rank. Now, because all solutions to (1) lie inside the unit square, this full-rank condition in such set is both necessary and sufficient to apply the Implicit Function Theorem to all π_1^*, π_2^* that solve (1). Therefore $\big(\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})\big)$ is a $\mathcal{C}^1$ function of $\boldsymbol{\theta}$ around a neighborhood of $\boldsymbol{Z}$ -in fact, by the IFT it inherits all smooth properties of $G_1(\cdot)$ and $G_2(\cdot)$-. In this case, the likelihood $\mathcal{F}(y_1, y_2 \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ is a $\mathcal{C}^1$ function of $\boldsymbol{\theta}$ around a neighborhood of $\boldsymbol{Z}$ and for all $\boldsymbol{X} \in \mathbb{R}^{k+2}$ because both $\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})$ and $\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})$ are $\mathcal{C}^1$ functions of $\boldsymbol{\theta}$ around a neighborhood of $\boldsymbol{Z}$ and for all $\boldsymbol{X} \in \mathbb{R}^{k+2}$.

Proof of Corollary to Lemma 3.3: If $\alpha_1\alpha_2 < 1/(\overline{g}_1\overline{g}_2)$. Then ($\star$) is satisfied for all $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$. Consequently, all the results of Lemma 3.3 hold everywhere in $\mathbb{S}(\boldsymbol{Z})$: $\big(\pi_1^*(\boldsymbol{Z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{Z}, \boldsymbol{\theta})\big)$ are unique for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and $\mathcal{F}(y_1, y_2 \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ exists for each $\boldsymbol{Z} \in \mathbb{S}(\boldsymbol{Z})$ and each $\boldsymbol{X} \in \mathbb{R}^k$. The implicit function theorem holds everywhere in $\mathbb{S}(\boldsymbol{Z})$ and therefore $\mathcal{F}(y_1, y_2 \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta})$ is a $\mathcal{C}^1$ function of $\boldsymbol{\theta}$ for all $\boldsymbol{X} \in \mathbb{R}^k$ and everywhere in $\mathbb{S}(\boldsymbol{Z})$.

A.2 Proof of Lemma 4.1

Assumptions (S1) and (S2) are sufficient to satisfy ($\widetilde{S1}$) and ($\widetilde{S2}$) respectively. From assumption (S3.2), the additional condition of Lemma 3.3:

$$\alpha_1\alpha_2 E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \mid \boldsymbol{Z} = \boldsymbol{z}\big] E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \mid \boldsymbol{Z} = \boldsymbol{z}\big] < 1 \ \forall(\pi_1, \pi_2) \in [0,1]^2$$

is satisfied everywhere in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, each $(\boldsymbol{z}, \boldsymbol{\theta}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ has a unique solution $pmb\pi^*(\boldsymbol{\theta}, \boldsymbol{z})$ to the equilibrium conditions (1). From Lemma 3.2, we know that $pmb\pi^*(\boldsymbol{\theta}, \boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{z}, \boldsymbol{\theta})$. From assumption (S1.3), $(\varepsilon_1, \varepsilon_2)$ have infinite support. Therefore, compactness of $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\Theta} \times [0,1]^2$ implies that there exists $\tau \in (0,1)$ such that $G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2) \in (\tau, 1-\tau)$ and $G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1) \in (\tau, 1-\tau)$ with probability one for all $(\pi_1, \pi_2) \in [0,1]^2$. Consequently, $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in (\tau, 1-\tau) \subset (0,1)^2$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Now, notice that the determinant of the Jacobian of the equilibrium system (1): $\nabla_{\boldsymbol{\pi}}\big(\boldsymbol{\pi} - \varphi(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)$ is given by $1-\alpha_1\alpha_2 E\big[g_1(\boldsymbol{X}_1'\boldsymbol{\beta}_1+\alpha_1\pi_2) \mid \boldsymbol{Z}=\boldsymbol{z}\big]E\big[g_2(\boldsymbol{X}_2'\boldsymbol{\beta}_2+\alpha_2\pi_1) \mid \boldsymbol{Z}=\boldsymbol{z}\big]$, which by assumption (S3.2) is nonzero (strictly positive) everywhere in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Consequently, the Implicit Function Theorem (IFT) holds for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\mathcal{Z}}$ and therefore $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ is a well-defined function of $\boldsymbol{\theta}$ and $\boldsymbol{z}$ everywhere in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and it inherits all the smoothness properties of $\varphi(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$. Therefore, using assumptions (S1.3) and (S2.3) we have that $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ is M times differentiable functions of $(\boldsymbol{\theta}, \boldsymbol{Z})$ with bounded M^{th} derivatives everywhere in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. In particular, let $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ and $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ be the matrices of first and second derivatives with respect to $\boldsymbol{\theta}$. Then, using the IFT we have:

$$\underset{2\times(k+2)}{\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})} = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$$

$$\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})} = \nabla_{\boldsymbol{\theta}}\text{vec}\Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\Big)$$

where $\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$ is the partial derivative of $\varphi(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ (with $\boldsymbol{\pi}$ fixed) evaluated at $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$. On the other hand, $\nabla_{\boldsymbol{\theta}}\text{vec}\Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\Big)$ includes $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$. □

A.3 Identification in the linear model

Recall that the equilibrium probabilities in the linear version of the model presented in section 4.4.1 are given by:

$$\pi_1^*(\boldsymbol{\theta}, \boldsymbol{z}) = \frac{2\big[E[\boldsymbol{X}_1 \mid \boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_1 + 1\big] + \alpha_1\big[E[\boldsymbol{X}_2 \mid \boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_2 + 1\big]}{4-\alpha_1\alpha_2}$$
$$\pi_2^*(\boldsymbol{\theta}, \boldsymbol{z}) = \frac{2\big[E[\boldsymbol{X}_2 \mid \boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_2 + 1\big] + \alpha_2\big[E[\boldsymbol{X}_1 \mid \boldsymbol{Z}=\boldsymbol{z}]'\boldsymbol{\beta}_1 + 1\big]}{4-\alpha_1\alpha_2}$$

where $4-\alpha_1\alpha_2 > 0$ by assumption (S3.2). Consequently, we can express:

$$\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{\theta}, \boldsymbol{Z}) = \delta_1 + \boldsymbol{X}_1'\boldsymbol{\beta}_1 + E[\boldsymbol{X}_1 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{1,1} + E[\boldsymbol{X}_2 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{1,2}$$

$$\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1^*(\boldsymbol{\theta}, \boldsymbol{Z}) = \delta_2 + \boldsymbol{X}_2'\boldsymbol{\beta}_2 + E[\boldsymbol{X}_1 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{2,1} + E[\boldsymbol{X}_2 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{2,2}$$

Let $d \equiv 4 - \alpha_1\alpha_2$. Suppose we allow for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ to include a constant term and denote the coefficients for these constants (i.e intercepts) by $\beta_{1,c}$ and $\beta_{2,c}$ respectively. Then:

$$\delta_1 = \frac{2\alpha_1 + \alpha_1\alpha_2 + 4\beta_{1,c} + 2\beta_{2,c}\alpha_1}{d}, \quad \delta_2 = \frac{2\alpha_2 + \alpha_1\alpha_2 + 4\beta_{2,c} + 2\beta_{1,c}\alpha_2}{d}$$

$$\gamma_{1,1} = \frac{\boldsymbol{\beta}_1\alpha_1\alpha_2}{d}, \quad \gamma_{1,2} = \frac{2\boldsymbol{\beta}_2\alpha_1}{d}, \quad \gamma_{2,1} = \frac{2\boldsymbol{\beta}_1\alpha_2}{d} \quad \text{and} \quad \gamma_{2,2} = \frac{\boldsymbol{\beta}_2\alpha_1\alpha_2}{d}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ exclude the intercepts $\beta_{1,c}$ and $\beta_{2,c}$, which are included in δ_1 and δ_2. As these functions show, we would be able to identify all the parameters (including $\beta_{1,c}$ and $\beta_{2,c}$) if we could recover α_1 and α_2. Suppose $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ have full-column rank and there exist $X_{1,\ell_1} \in \boldsymbol{X}_1$ and $X_{2,\ell_2} \in \boldsymbol{X}_2$ such that $\beta_{1,\ell_1} \neq 0$, $\beta_{2,\ell_2} \neq 0$, $E[X_{1,\ell_1} \mid \boldsymbol{Z}] \neq X_{1,\ell_1}$ and $E[X_{2,\ell_2} \mid \boldsymbol{Z}] \neq X_{2,\ell_2}$. We next show how to recover α_1 and α_2:

- If $\gamma_{2,1_{\ell_1}} \neq 0$ and $\gamma_{1,2_{\ell_2}} \neq 0$ then $\alpha_1 = \gamma_{1,2_{\ell_2}}/(2\gamma_{2,2_{\ell_2}})$ and $\alpha_2 = \gamma_{2,1_{\ell_1}}/(2\gamma_{1,1_{\ell_1}})$

- If $\gamma_{2,1_{\ell_1}} \neq 0$ and $\gamma_{1,2_{\ell_2}} = 0$ then $\alpha_1 = 0$ and $\alpha_2 = 2\gamma_{2,1_{\ell_1}}/\beta_{1,\ell_1}$

- If $\gamma_{2,1_{\ell_1}} = 0$ and $\gamma_{1,2_{\ell_2}} \neq 0$ then $\alpha_1 = 2\gamma_{1,2_{\ell_2}}/\beta_{2,\ell_2}$ and $\alpha_2 = 0$

- If $\gamma_{2,1_{\ell_1}} = 0$ and $\gamma_{1,2_{\ell_2}} = 0$ then $\alpha_1 = 0$ and $\alpha_2 = 0$.

we use α_1 and α_2 to recover the intercepts $\beta_{1,c}$ and $\beta_{2,c}$ as follows:

$$\beta_{1,c} = \frac{2\delta_1 - \alpha_1(1+\delta_2)}{2} \quad \text{and} \quad \beta_{2,c} = \frac{2\delta_2 - \alpha_2(1+\delta_1)}{2}$$

now suppose there exists $X_{1_{\kappa_1}} \in \boldsymbol{X}_1$ such that $E[X_{1,\kappa_1} \mid \boldsymbol{Z}] = X_{1,\kappa_1}$. Then we would have $X_{1_{\kappa_1}}\beta_{1_{\kappa_1}} + E[X_{1,\kappa_1} \mid \boldsymbol{Z}]\gamma_{1_{\kappa_1}} = X_{1_{\kappa_1}}\beta_{1_{\kappa_1}}(1 + \alpha_1\alpha_2/d) = 4X_{1_{\kappa_1}}\beta_{1_{\kappa_1}}/d$, which clearly shows we can recover $\beta_{1_{\kappa_1}}$ by excluding $E[X_{1,\kappa_1} \mid \boldsymbol{Z}]$ and including only X_{1,κ_1} in the equation $\delta_1 + \boldsymbol{X}_1'\boldsymbol{\beta}_1 + E[\boldsymbol{X}_1 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{1,1} + E[\boldsymbol{X}_2 \mid \boldsymbol{Z}]'\boldsymbol{\gamma}_{1,2}$. Let $\widetilde{\beta}_{1_{\kappa_1}}$ denote the corresponding coefficient, then we have $\beta_{1_{\kappa_1}} = (d/4)\widetilde{\beta}_{1_{\kappa_1}}$. We would follow parallel steps to recover the coefficient $\beta_{1_{\kappa_2}}$ for any $X_{2_{\kappa_2}} \in \boldsymbol{X}_2$ such that $E[X_{2,\kappa_2} \mid \boldsymbol{Z}] = X_{2,\kappa_2}$.

Now suppose $E[\boldsymbol{X}_1 \mid \boldsymbol{Z}] = \boldsymbol{X}_1$ and $E[\boldsymbol{X}_2 \mid \boldsymbol{Z}] = \boldsymbol{X}_2$. Then we get:

$$\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\pi_2^*(\boldsymbol{\theta}, \boldsymbol{Z}) = \delta_1 + \boldsymbol{X}_1'\left(\frac{4\boldsymbol{\beta}_1}{d}\right) + \boldsymbol{X}_2'\left(\frac{2\boldsymbol{\beta}_2\alpha_1}{d}\right) \equiv \delta_1 + \boldsymbol{X}_1'\widetilde{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2'\widetilde{\gamma}_{1,2}$$

$$\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\pi_1^*(\boldsymbol{\theta}, \boldsymbol{Z}) = \delta_2 + \boldsymbol{X}_2'\left(\frac{4\boldsymbol{\beta}_2}{d}\right) + \boldsymbol{X}_1'\left(\frac{2\boldsymbol{\beta}_1\alpha_2}{d}\right) \equiv \delta_2 + \boldsymbol{X}_2'\widetilde{\boldsymbol{\beta}}_2 + \boldsymbol{X}_1'\widetilde{\gamma}_{2,1}$$

where δ_1 and δ_2 are as defined above. Now suppose there exist $X_{1,\ell_1} \in \boldsymbol{X}_1$ and $X_{2,\ell_2} \in \boldsymbol{X}_2$ such that $\beta_{1,\ell_1} \neq 0$ and $\beta_{2,\ell_2} \neq 0$. Then it is easy to see that $\alpha_1 = 2\widetilde{\gamma}_{1,2_{\ell_2}}/\widetilde{\beta}_{2,\ell_2}$ and $\alpha_2 = 2\widetilde{\gamma}_{2,1_{\ell_1}}/\widetilde{\beta}_{1,\ell_1}$. The intercepts $\beta_{1,c}$ and $\beta_{2,c}$ would be recovered in the same way as it was described above and we would trivially recover the slope parameters by $\boldsymbol{\beta}_1 = (d/4) * \widetilde{\boldsymbol{\beta}}_1$ and $\boldsymbol{\beta}_2 = (d/4) * \widetilde{\boldsymbol{\beta}}_2$. If no such $X_{1,\ell_1} \in \boldsymbol{X}_1$ and $X_{2,\ell_2} \in \boldsymbol{X}_2$ exist, then it is not possible to identify the intercepts $\beta_{1,c}$ and $\beta_{2,c}$ along with the strategic parameters α_1, α_2. In this case, if we normalize $\beta_{1,c} = 0$ and $\beta_{2,c} = 0$ then we would have $\alpha_1 = (2\delta_1)/(1+\delta_2)$ and $\alpha_2 = (2\delta_2)/(1+\delta_1)$.

A.4 Lemmas 4.4, 4.5

Proof of Lemma 4.4:

We have

$$\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)} = \left\{ \frac{\Pr(\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})}{\Pr(\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)} \quad \text{if } \boldsymbol{Z} \in \boldsymbol{\mathcal{Z}} \quad \text{and} \quad 1 \quad \text{otherwise} \right\}$$

where $\boldsymbol{Y} \in \{(1,1),(1,0),(0,1),(0,0)\}$ and $\Pr(\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ is the conditional probability of $\boldsymbol{Y}$ given $(\boldsymbol{X},\boldsymbol{Z})$ when the parameter vector equals $\boldsymbol{\theta}$. If assumption (S5) is satisfied, then $\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}$ is not constant whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Note also that by definition this ratio is always positive for every $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Therefore, by Jensen's inequality we have:

$$-\log\left\{E\left[\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right]\right\} < E\left[-\log\left\{\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right\}\right]$$

If assumptions (I), (S1.1-2), (S2.1-2) and (S3.2) are satisfied then if $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$, we have:

$$\Pr\left\{\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)} = \frac{\Pr(\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})}{\Pr(\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)}\right\} = \Pr(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)$$

for each $\boldsymbol{y} \in \{(1,1),(1,0),(0,1),(0,0)\}$. Therefore:

$$\begin{aligned}
E\left[\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right] &= \int_{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})}\left\{\sum_{\boldsymbol{y}}\frac{\Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})}{\Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)}\cdot\Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)\right\}f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z} \\
&\quad + \int_{\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})/\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})} 1\cdot f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z} \\
&= \int_{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})}\left\{\sum_{\boldsymbol{y}}\Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})\right\}f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z} + (1-\Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}})) \\
&= \Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}) + (1-\Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}})) = 1
\end{aligned}$$

where the last equality uses the fact that $\sum_{\boldsymbol{y}} \Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}) = 1$ for all $(\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Z})$. Therefore we get that whenever $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$:

$$0 < E\left[-\log\left\{\frac{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right\}\right] = E[\log \mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)] - E[\log \mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})]$$

or equivalently: $E\big[log\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})\big] < E\big[log\mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)\big] \quad \forall\, \boldsymbol{\theta} \neq \boldsymbol{\theta}_0,\ \boldsymbol{\theta} \in \boldsymbol{\Theta}$, which proves the claim. □

Proof of Lemma 4.5:

First, recall that:

$$\widetilde{\mathcal{F}}(\boldsymbol{W},\boldsymbol{\theta}) = G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\rho_2(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_1}\big[1 - G_1\big(\boldsymbol{X}_1'\boldsymbol{\beta}_1 + \alpha_1\rho_2(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_1}$$
$$\times\, G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\rho_1(\boldsymbol{\theta},\boldsymbol{Z})\big)^{Y_2}\big[1 - G_2\big(\boldsymbol{X}_2'\boldsymbol{\beta}_2 + \alpha_2\rho_1(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]^{1-Y_2}$$

with $\rho(\boldsymbol{\theta},\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\big]$. The proof follows basically the same steps as that of Lemma 4.4. If (S1.1-2) and (S2.1-2) are satisfied, then assumption (S3.2) precludes the situation $\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Therefore, if (S5) is also satisfied we have that conditional on $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$, if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ with $\boldsymbol{\theta},\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ then:

$$\Pr\Big\{\boldsymbol{\beta}_1'\boldsymbol{X}_1 + \alpha_1\rho_2(\boldsymbol{\theta},\boldsymbol{Z}) \neq \boldsymbol{\beta}_{1_0}'\boldsymbol{X}_1 + \alpha_{1_0}\pi_2^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\Big\} > 0$$
$$\Pr\Big\{\boldsymbol{\beta}_2'\boldsymbol{X}_2 + \alpha_2\rho_1(\boldsymbol{\theta},\boldsymbol{Z}) \neq \boldsymbol{\beta}_{2_0}'\boldsymbol{X}_2 + \alpha_{2_0}\pi_1^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\Big\} > 0$$

Therefore $\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})/\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)$ is not constant in $\boldsymbol{\mathcal{Z}}$. It is also everywhere positive and therefore the same Jensen's inequality argument used in the proof of Lemma 4.4 applies:

$$-\log\left\{E\left[\frac{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right]\right\} < E\left[-\log\left\{\frac{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right\}\right]$$

Now recall that $\widetilde{\mathcal{F}}(\boldsymbol{W},\boldsymbol{\theta}_0) = \mathcal{F}(\boldsymbol{W},\boldsymbol{\theta}_0)$ and $\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0) = \mathcal{F}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)$ (the true likelihood and trimmed likelihood respectively) everywhere in $\boldsymbol{\mathcal{Z}}$. Therefore, if assumptions (I), (S1.1-2), (S2.1-2) and (S3.2) are satisfied then if $\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}$, we have:

$$\Pr\left\{\frac{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)} = \frac{\widetilde{\mathcal{F}}(\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})}{\Pr(\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)}\right\} = \Pr(\boldsymbol{Y}=\boldsymbol{y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)$$

for each $\boldsymbol{y} \in \{(1,1),(1,0),(0,1),(0,0)\}$.

Therefore:

$$
\begin{aligned}
E\left[\frac{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right] =& \int_{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})}\left\{\sum_{\boldsymbol{y}}\frac{\widetilde{\mathcal{F}}(\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})}{\Pr(\boldsymbol{Y}=\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)}\cdot\Pr(\boldsymbol{Y}=\boldsymbol{y}\mid\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)\right\}f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z} \\
&+\int_{\boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z})/\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})}1\cdot f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z} \\
=&\int_{\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}}\int_{\boldsymbol{X}\in\mathbb{S}(\boldsymbol{X})}\left\{\sum_{\boldsymbol{y}}\widetilde{\mathcal{F}}(\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})\right\}f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{X},\boldsymbol{Z})d\boldsymbol{X}d\boldsymbol{Z}+(1-\Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}})) \\
=&\Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}})+(1-\Pr(\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}))=1
\end{aligned}
$$

where the last equality uses the fact that $\sum_{\boldsymbol{y}}\widetilde{\mathcal{F}}(\boldsymbol{y},\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})=1$ for all $(\boldsymbol{\theta},\boldsymbol{X},\boldsymbol{Z})$. Therefore we get that whenever $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$:

$$
0<E\left[-\log\left\{\frac{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})}{\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)}\right\}\right]=E[\log\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)]-E[\log\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})]
$$

or equivalently: $E\big[log\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta})\big]<E\big[log\widetilde{\mathcal{F}}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W},\boldsymbol{\theta}_0)\big]\quad\forall\,\boldsymbol{\theta}\neq\boldsymbol{\theta}_0,\ \boldsymbol{\theta}\in\boldsymbol{\Theta}$, which proves the claim. □

A.5 Theorems 1, 2

We first need to establish the uniform rate of convergence of the proposed estimators for $\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta})$. The next lemma is an application of Lemma 3 in Collomb and Hardle (1986). Variants of the latter result have been used previously by Stoker (1991) and Ahn and Manski (1993).

Lemma A.1 *Let $\{(\boldsymbol{X}_n,\boldsymbol{Z}_n)\}_{n=1}^N$ be an iid sequence in $\mathbb{R}^K\times\mathbb{R}^L$, with $\boldsymbol{X}_n$ bounded with probability one. Suppose we have a kernel $K:\mathbb{R}^L\to\mathbb{R}$ that is symmetric, bounded and satisfies the conditions: $\|u\|\cdot|K(u)|\to 0$ as $\|u\|\to\infty$, $\int K(u)du=1$ and the Lipschitz condition: $\exists\gamma>0$, $c_k<\infty$ such that $|\ K(u)-K(v)\ |\leq c_k\|u-v\|^\gamma\ \forall u,v\in\mathbb{R}^L$. Suppose the sequence $\{h_N;N\in\mathbb{N}\}$ is such that as $N\to\infty$: $h_N\to 0$ and $Nh_N^L/log\,N\to\infty$. Let $\eta:\mathbb{R}^K\times\mathbb{R}^L\times\mathbb{R}^P\to\mathbb{R}$ be a continuously differentiable function that satisfies: $\big|\eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})\big|\leq$*

$\overline{M} < \infty$, $\left\| \frac{\partial \eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})}{\partial \boldsymbol{t}} \right\| \leq \overline{C}_1 < \infty$ *and* $\left\| \frac{\partial \eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})}{\partial \boldsymbol{z}} \right\| \leq \overline{C}_2 < \infty$ *for all* $(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})$. *Now let:*

$$R_N(\boldsymbol{z},\boldsymbol{t}) = \frac{1}{Nh_N^L} \sum_{n=1}^{N} \eta(\boldsymbol{X}_n,\boldsymbol{z},\boldsymbol{t}) K\left(\frac{\boldsymbol{Z}_n - \boldsymbol{z}}{h_N} \right)$$

Then, for any compact sets $\boldsymbol{C} \in \mathbb{R}^L$ *and* $\boldsymbol{G} \in \mathbb{R}^P$ *and any* $\varepsilon > 0$ *we have:*

$$\left(N^{1-\varepsilon} h_N^L\right)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{t}\in\boldsymbol{G}}} \left| R_N(\boldsymbol{z},\boldsymbol{t}) - ER_N(\boldsymbol{z},\boldsymbol{t}) \right| = O_p(1) \quad \textit{w.p.1}$$

Proof: If the assumptions outlined above are satisfied, then using Lemma 3 in Collomb and Hardle, we have that for every compact sets $\boldsymbol{C} \in \mathbb{R}^L$ and $\boldsymbol{G} \in \mathbb{R}^P$:

$$\left(Nh_N^L/\log N\right)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{t}\in\boldsymbol{G}}} \mid R_N(\boldsymbol{z},\boldsymbol{t}) - ER_N(\boldsymbol{z},\boldsymbol{t}) \mid = O_p(1) \quad \text{w.p.1}$$

Therefore:

$$\left(N^{1-\varepsilon} h_N^L\right)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{t}\in\boldsymbol{G}}} \mid R_N(\boldsymbol{z},\boldsymbol{t}) - ER_N(\boldsymbol{z},\boldsymbol{t}) \mid = \left(\frac{\log N}{N^{\varepsilon}} \right)^{1/2} O_p(1) = O_p(1) \quad \text{for all } \varepsilon > 0 \quad \text{w.p.1}$$

which shows the result. □

Lemma A.1 is sufficient to show the results that follow, which rely on a weaker version of it. Before proceeding, let us present the following notation: We will let $p \in \{1,2\}$ and define $-p$ as: $-p = 2$ if $p = 1$ and $-p = 1$ if $p = 2$. Now take $\boldsymbol{\theta} \in \mathbb{R}^{K+2}$ and $\boldsymbol{z} \in \mathbb{S}(\boldsymbol{Z})$. Following conventional notation, let $g_p^{(m)}(\cdot)$ represent the m^{th} derivative of $g_p(\cdot)$. Then, for $p \in \{1,2\}$ and $\pi_{-p} \in \mathbb{R}$ define:

$$\begin{aligned}
\varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) =& E\left[G_p(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right] \\
\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) =& E\left[g_p(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right] \\
\delta_p^{(m)}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) =& E\left[g_p^{(m)}(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right] \quad \text{with } m \geq 1 \\
\zeta_p^{(m)}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) =& E\left[\boldsymbol{X}_p \; g_p^{(m)}(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right] \quad \text{with } m \geq 0 \\
\xi_p^{(m)}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) =& E\left[\left(\boldsymbol{X}_p\boldsymbol{X}_p'\right) g_p^{(m)}(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right] \quad \text{with } m \geq 0
\end{aligned}$$

the following result is a consequence of Lemma A.1 and assumptions (S1.3), (S2) and (S4).

Lemma A.2 *Suppose assumptions (S1.3), (S2) and (S4) are satisfied. Let:*

$$\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

and for $p \in \{1,2\}$ define:

$$\widehat{\varphi}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{G_p\big(\boldsymbol{X'}_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$$

$$\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{g_p\big(\boldsymbol{X'}_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$$

$$\widehat{\delta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{g_p^{(m)}\big(\boldsymbol{X'}_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \textit{with } m \geq 1$$

$$\widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{\boldsymbol{X}_{p_n} g_p^{(m)}\big(\boldsymbol{X'}_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \textit{with } m \geq 0$$

$$\widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \frac{\big(\boldsymbol{X}_{p_n}\boldsymbol{X'}_{p_n}\big) g_p^{(m)}\big(\boldsymbol{X'}_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \textit{with } m \geq 0$$

Let $\boldsymbol{C}$ be any compact set in the interior of $\mathbb{S}(\boldsymbol{Z})$ such that $\inf_{\boldsymbol{z}\in\boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$. Then

$$\text{(A)} \qquad \sup_{\boldsymbol{z}\in\boldsymbol{C}} \left|\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\right| = o_p(N^{-1/4})$$

Now take any compact sets $\boldsymbol{A} \in \mathbb{R}$ and $\boldsymbol{B} \in \mathbb{R}^{k_p+1}$. Then, for $p \in \{1,2\}$ we have:

$$\text{(B)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{\varphi}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \varphi_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right| = o_p(N^{-1/4})$$

$$\text{(C)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \delta_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right| = o_p(N^{-1/4})$$

$$\text{(D)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{\delta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \delta^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right| = o_p(N^{-1/4}) \quad m = 1, \ldots, M+1$$

$$\text{(E)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left\|\widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \zeta^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right\| = o_p(N^{-1/4}) \quad m = 0, \ldots, M+1$$

$$\text{(F)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left\|\widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \xi^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right\| = o_p(N^{-1/4}) \quad m = 0, \ldots, M+1$$

Proof:

To show (A), we first prove that there exists $\widetilde{D}_1 < \infty$ such that

$$\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| \leq \widetilde{D}_1 \cdot h_N^M$$

Define:

$$Q_i \equiv \{(q_1,\ldots,q_L) \in \mathbb{N}^L : q_1 + \cdots + q_L = i\} \quad \text{and} \quad \Gamma_i(\boldsymbol{z}) = \sum_{Q_i} \frac{\partial^i f_{\boldsymbol{Z}}(\boldsymbol{z})}{\partial z_1^{q_1}\cdots \partial z_L^{q_L}}$$

then by (S2.2) the following Taylor series approximation is valid:

$$E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) = \int K(\boldsymbol{\Psi}) f_{\boldsymbol{Z}}(\boldsymbol{z} + h_N \boldsymbol{\Psi}) d\boldsymbol{\Psi} = f_{\boldsymbol{Z}}(\boldsymbol{z}) \int K(\boldsymbol{\Psi}) d\boldsymbol{\Psi} + \sum_{i=1}^{M-1} (-1)^i \frac{h_N^i}{i!} \Gamma_i(\boldsymbol{z}) \sum_{Q_i} \int \Psi_1^{q_1}\cdots\Psi_L^{q_L} K(\boldsymbol{\Psi}) d\boldsymbol{\Psi}$$
$$+ (-1)^M \frac{h_N^M}{M!} \int \sum_{Q_M} (\Psi_1^{q_1}\cdots\Psi_L^{q_L}) \Gamma_M(\boldsymbol{z} + h_N^* \boldsymbol{\Psi}) K(\boldsymbol{\Psi}) d\boldsymbol{\Psi}$$

where h_N^* is between h_N and zero. By (S2.2) there exists a $\overline{C}_1 < \infty$ such that $\left|\Gamma_i(\boldsymbol{v})\right| < \overline{C}_1$ for all $\boldsymbol{v} \in \mathbb{R}^L$ and all $i \in \{1,\ldots,M\}$. This, along with (S4.1) implies that

$$\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| \leq \widetilde{D}_1 h_N^M \quad \text{where} \quad \widetilde{D}_1 = \frac{1}{M!} \overline{C}_1 \left|Q_M\right| \int \left\|\boldsymbol{\Psi}\right\|^M d\boldsymbol{\Psi}$$

where $\left|Q_M\right|$ denotes the number of elements in the set Q_M. Take ε described in (S4.2(i)) and let $\eta(\cdot,\cdot,\cdot) = 1$, then by (S2.2), (S4.1(i)-(iii)), (S4.2(i)), (S4.3) and the compactness of $\boldsymbol{C}$, all the assumptions of lemma A.1 are satisfied[21] and we get:

$$\left(N^{1-\varepsilon} h_N^L\right)^{1/2} \sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \right| = O_P(1)$$

Using the inequality $\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| \leq \sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \right| + \sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| f_{\boldsymbol{Z}}(\boldsymbol{z}) - E\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \right|$, we then have:

$$N^{1/4} \sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| \leq \left(N^{1-2\varepsilon} h_N^{2L}\right)^{-1/4} O_p(1) + N^{1/4} h_N^M \widetilde{D}_1 = o_p(1)$$

where the last equality follows from (S4.2(i-ii)). Equivalently: $\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| = o_p(\mathrm{N}^{-1/4})$, which establishes (A).

[21] Note that $N^{1-2\varepsilon} h_N^{2L} \to \infty \Rightarrow N^{1-\varepsilon} h_N^L \to \infty \Rightarrow N h_N^L / \log \mathrm{N} \to \infty$.

To prove results (B)-(E), note first that we can express

$$\widehat{\varphi}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}, \quad \widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$$

$$\widehat{\delta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}, \quad \widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{T}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$$

$$\widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{t}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$$

where

$$\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} G_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} g_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} g_p^{(m)}\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\widehat{T}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \boldsymbol{X}_{p_n}\; g_p^{(m)}\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\widehat{t}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} \big(\boldsymbol{X}_{p_n}\boldsymbol{X}'_{p_n}\big) g_p^{(m)}\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

We begin by examining $\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$. We will proceed in a similar fashion as in part (A), and show that there exists $\widetilde{D}_2 < \infty$ such that

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \Big| E\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big| \le \widetilde{D}_2 \cdot h_N^M$$

Take Q_i to be the set defined above and let:

$$\Gamma_i^p(\boldsymbol{z}) = \sum_{Q_i} \frac{\partial^i f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{x}_p,\boldsymbol{z})}{\partial z_1^{q_1}\cdots\partial z_L^{q_L}}$$

then by (S2.2) the following Taylor series approximation is valid:

$$f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}+h_N\boldsymbol{\Psi}) = f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}) + \sum_{i=1}^{M-1}(-1)^i\frac{h_N^i}{i!}\Gamma_i^p(\boldsymbol{u},\boldsymbol{z})\sum_{Q_i}\Psi_1^{q_1}\cdots\Psi_L^{q_L} + (-1)^M\frac{h_N^M}{M!}\sum_{Q_M}(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma_M^p(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi}) \tag{4}$$

therefore we have:

$$E\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \int\int G_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})K(\boldsymbol{\Psi})f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z} + h_N\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u} =$$

$$\int G_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})\Bigg\{f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z})\int K(\boldsymbol{\Psi})d\boldsymbol{\Psi} + \sum_{i=1}^{M-1}(-1)^i\frac{h_N^i}{i!}\Gamma_i^p(\boldsymbol{u},\boldsymbol{z})\sum_{Q_i}\int \Psi_1^{q_1}\cdots\Psi_L^{q_L}K(\boldsymbol{\Psi})d\boldsymbol{\Psi}$$

$$+ (-1)^M\frac{h_N^M}{M!}\int\sum_{Q_M}(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma_M^p(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}\Bigg\}d\boldsymbol{u}$$

where h_N^* is between h_N and zero. Now, because $\boldsymbol{C}$ is a compact set in the interior of $\mathbb{S}(\boldsymbol{Z})$ and $\inf_{\boldsymbol{z}\in\boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$, which means that:

$$\int G_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})f_{\boldsymbol{X}_p,\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z})d\boldsymbol{u} = f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) \quad \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$$

By (S2.2) there exists a $\overline{C}_2 < \infty$ such that $\left|\Gamma_i^p(\boldsymbol{u},\boldsymbol{v})\right| < \overline{C}_2$ for all $(\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}^{k_p}\times\mathbb{R}^L$ and all $i \in \{1,\ldots,M\}$. We also have $G_p(v) \in (0,1)\ \forall\ v \in \mathbb{R}$. These results, along with (S4.1) and the approximation described above implies that for all $(\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$ we have:

$$\left|E\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| =$$

$$\frac{h_N^M}{M!}\left|\int\int\sum_{Q_M}G_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma_M^p(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u}\right|$$

$$\leq \widetilde{D}_2\cdot h_N^M \quad \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$$

with $\widetilde{D}_2 = \frac{1}{M!}\overline{C}_2\left|Q_M\right|\int\left\|\boldsymbol{\Psi}\right\|^M d\boldsymbol{\Psi}$ where, as before, $\left|Q_M\right|$ represents the number of elements of the set Q_M.

Now define $\boldsymbol{t} = (\boldsymbol{\theta}_p,\pi_{-p})$ and let $\eta(X_p,\cdot,\boldsymbol{t}) = G_p\big(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)$. Take ε as described in (S4.2(i)). Recall that $G_p(v) \in (0,1)$ for all $v \in \mathbb{R}$. This, along with (S2.2-3), (S4.1(i)-(iii)), (S4.2(i)), (S4.3) and the compactness of $\boldsymbol{C},\boldsymbol{B}$ and $\boldsymbol{A}$ implies that all the assumptions of Lemma A.1 are satisfied. Therefore:

$$\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}}\left|\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - E\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1)$$

Using the inequality

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - E\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| + \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|E\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right|$$

we then have:

$$N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq \left(N^{1-2\varepsilon}h_N^{2L}\right)^{-1/4}O_p(1) + N^{1/4}h_N^{M}\widetilde{D}_2$$

and by (S4.2(i-ii)) we get:

$$N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = o_p(1) \qquad (1\star)$$

We have $\widehat{\varphi}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$. Therefore by a first-order Taylor approximation we get:

$$\widehat{\varphi}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) = \varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) + \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}\Big[\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big] - \frac{\widetilde{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^2}\Big[\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\Big]$$

with $\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ between $\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $\widetilde{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$ between $\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$. By (A) and the fact that $\inf_{\boldsymbol{z}\in\boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$, we have:

$$\sup_{\boldsymbol{z}\in\boldsymbol{C}}\left|\frac{1}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}\right| = O_p(1) \quad \text{and therefore} \quad \sup_{\boldsymbol{z}\in\boldsymbol{C}}\left|\frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}\right| = O_p(1)$$

By $(1\star)$ and the fact that $\varphi_p(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\in(0,1)\ \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p})\in\mathbb{S}(\boldsymbol{Z})\times\mathbb{R}^{k_p+1}\times\mathbb{R}$, we also have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left|\widetilde{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1)$$

Therefore:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}} \left|\widehat{\varphi}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq O_p(1)\cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}} \left|\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right|$$

$$+ O_p(1)\cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}} \left|\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\right|$$

$$= o_p(1)$$

with the last equality following from (A) and (1$\star$). This proves part (B) of the lemma.

Now take $\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$. Using (4) and the same arguments as above, we can show that for all $(\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$ we have:

$$\left|E\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| =$$
$$\frac{h_N^M}{M!}\left|\int\int\sum_{Q_M} g_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma_M^p(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u}\right|$$

Now let $\overline{g}_p = \mathrm{Max}_{v\in\mathbb{R}} g_p(v)$. By (S1.3) we have $\overline{g}_p < \infty$ and therefore:

$$\left|E\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq \widetilde{D}_3 h_N^M \quad \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$$

with $\widetilde{D}_3 = \frac{1}{M!}\overline{C}_2\left|Q_M\right|\overline{g}_p\int\left\|\boldsymbol{\Psi}\right\|^M d\boldsymbol{\Psi}$ where $\overline{C}_2$ and $\left|Q_M\right|$ are as defined above.

As before, define $\boldsymbol{t} = (\boldsymbol{\theta}_p,\pi_{-p})$ and now let $\eta(X_p,\cdot,\boldsymbol{t}) = g_p\big(\boldsymbol{X}_p'\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)$. Take ε as described in (S4.2(i)). Recall that $g_p(v) \in (0,\overline{g}_p]$ for all $v\in\mathbb{R}$. This, along with (S2.2-3), (S4.1(i)-(iii)), (S4.2(i)), (S4.3) and the compactness of $\boldsymbol{C},\boldsymbol{B}$ and $\boldsymbol{A}$ implies that all the assumptions of Lemma A.1 are satisfied. Therefore:

$$\big(N^{1-\varepsilon}h_N^L\big)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}} \left|\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - E\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1)$$

once again, using these results along with the triangle inequality we get:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}} \left|\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq \big(N^{1-2\varepsilon}h_N^{2L}\big)^{-1/4}O_p(1) + N^{1/4}h_N^M\widetilde{D}_3$$

and using (S4.2(i-ii)) we have:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = o_p(1) \qquad (2\star)$$

We have $\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{\widehat{s}_{p_N}(\pi_{-p}|\boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$. Therefore by a first-order Taylor approximation we get:

$$\begin{aligned}\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) &+ \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}\left[\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right] \\ &- \frac{\widetilde{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^2}\left[\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\right]\end{aligned}$$

with $\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ between $\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $\widetilde{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ between $\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$. We established above that by (A) and the fact that $\inf_{\boldsymbol{z}\in\boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$, we have:

$$\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left|\frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}\right| = O_p(1)$$

By $(2\star)$ and the fact that $\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) \in (0,\overline{g}_p]\ \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \mathbb{S}(\boldsymbol{Z})\times\mathbb{R}^{k_p+1}\times\mathbb{R}$, we also have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widetilde{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = O_p(1)$$

Therefore:

$$\begin{aligned} N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{\delta}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| &\leq O_p(1)\cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \\ &+ O_p(1)\cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{-p}\in\boldsymbol{A}}} \left|\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\right| \\ &= o_p(1) \end{aligned}$$

with the last equality following from (A) and $(2\star)$. This proves part (C) of the lemma.

The analysis of $\widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ is virtually identical to that of $\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$: we know that assumption (S1.3) implies that for $p \in \{1,2\}$, there exists $\overline{g}_p$ such that $g_p(v) < \overline{g}_p$ for all $v \in \mathbb{R}$. It also implies that for $p \in \{1,2\}$, there also exists $\overline{g}'_p < \infty$ such that $\left|g_p(v)^{(m)}\right| < \overline{g}'_p$

for all $v \in \mathbb{R}$ and all $m = 1, \ldots, M+1$ [22]. Therefore, following the same steps as above we can show that:

$$\left| E\widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| \leq \widetilde{D}_4 h_N^M \quad \forall\ (\boldsymbol{z}, \boldsymbol{\theta}_p, \pi_{-p}) \in \boldsymbol{C} \times \boldsymbol{B} \times A$$
$$\forall\ m = 1, \ldots, M+1$$

with $\widetilde{D}_4 = \frac{1}{M!}\overline{C}_2 \left|Q_M\right| \overline{g}'_p \int \left\|\boldsymbol{\Psi}\right\|^M d\boldsymbol{\Psi}$ where $\overline{C}_2$ and $\left|Q_M\right|$ are as defined above. As before, define $\boldsymbol{t} = (\boldsymbol{\theta}_p, \pi_{-p})$ and now let $\eta(X_p, \cdot, \boldsymbol{t}) = g^{(m)}_p\left(\boldsymbol{X}'_p\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\right)$. Take ε as described in (S4.2(i)). Recall that $\left|g_p(v)^{(m)}\right| < \overline{g}'_p$ for all $v \in \mathbb{R}$ and all $m = 1, \ldots, M+1$. This, along with (S2.2-3), (S4.1(i)-(iii)), (S4.2(i)), (S4.3) and the compactness of $\boldsymbol{C}, \boldsymbol{B}$ and $\boldsymbol{A}$ implies that all the assumptions of Lemma A.1 are satisfied for $m = 1, \ldots, M+1$. Therefore:

$$\left(N^{1-\varepsilon}h_N^L\right)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left| \widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - E\widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \quad \forall m = 1, \ldots, M+1$$

as before, using these results along with the triangle inequality we get:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left| \widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| \leq \left(N^{1-2\varepsilon}h_N^{2L}\right)^{-1/4} O_p(1) + N^{1/4}h_N^M \widetilde{D}_3$$

$\forall m = 1, \ldots, M+1$ and using (S4.2(i-ii)) we have:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p}\,\in\,\boldsymbol{A}}} \left| \widehat{s}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\delta^{(m)}_p(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = o_p(1) \quad \forall\ m = 1, \ldots, M+1 \tag{3$\star$}$$

To analyze $\widehat{T}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$, let X_{p_j} be the j^{th} element of $\boldsymbol{X}_p$, with $j \in \{1, \ldots, k_P\}$. Similarly, let:

$$\zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = E\left[X_{p_j} g^{(m)}_p(\boldsymbol{X}'_p\boldsymbol{\beta}_p + \alpha_p\pi_{-p}) \mid \boldsymbol{Z} = \boldsymbol{z}\right]$$
$$\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} X_{p_j,n}\ g^{(m)}_p\left(\boldsymbol{X}'_{p_n}\ \boldsymbol{\beta}_p + \alpha_p\pi_{-p}\right) K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

and note that by definition, we have $\zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \left(\zeta^{(m)}_{p_1}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p), \ldots, \zeta^{(m)}_{p_{k_p}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right)'$ and $\widehat{T}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \left(\widehat{T}^{(m)}_{p_{N_1}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p), \ldots, \widehat{T}^{(m)}_{p_{N_{k_p}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right)'$. Take any $j \in \{1, \ldots, k_p\}$,

[22] These results hold for $m = 0, \ldots, M+1$, but we will focus here on $m \geq 1$ since the case $m = 0$ corresponds to $\widehat{s}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$, which was analyzed in the previous paragraphs.

then once again using (4) and the same arguments used in the previous cases, we can show that for all $(\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$ we have:

$$\Big|E\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\zeta^{(m)}_{p_j}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big| = \frac{h_N^M}{M!}\Big|\int\int\sum_{Q_M} u_j g_p^{(m)}(\boldsymbol{u}'\boldsymbol{\beta}_p+\alpha_p\pi_{-p})(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma^p_M(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u}\Big|$$

By (S2.3), we have that $\mathbb{S}(\boldsymbol{X}_p)$ is a compact set, which means that $\exists\ \overline{X}<\infty$ such that $\mid X_{p_j}\mid<\overline{X}$ w.p.1 for all $j\in\{1,\ldots,k_p\}$. For $p\in\{1,2\}$ define $\overline{\kappa}_p = \text{Max}\ \{\overline{g}_p,\overline{g}'_p\}$. Then, we have [23]:

$$\Big|E\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\zeta_{p_j}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big| \le \widetilde{D}_5 h_N^M \quad \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p})\in\boldsymbol{C}\times\boldsymbol{B}\times A,\ \forall\ j\in\{1,\ldots,k_p\}\quad \forall\ m=0,\ldots M+1$$

with $\widetilde{D}_5 = \frac{1}{M!}\overline{C}_2\big|Q_M\big|\overline{\kappa}_p\overline{X}\int\big\|\boldsymbol{\Psi}\big\|^M d\boldsymbol{\Psi}$ where $\overline{C}_2$ and $\big|Q_M\big|$ are as defined above. As we have done in the previous cases, define $\boldsymbol{t}=(\boldsymbol{\theta}_p,\pi_{-p})$ and now let $\eta(X_p,\cdot,\boldsymbol{t}) = X_{p_j}g_p^{(m)}\big(\boldsymbol{X}'_p\boldsymbol{\beta}_p+\alpha_p\pi_{-p}\big)$. Take ε as described in (S4.2(i)). Recall that $\big|g_p(v)^{(m)}\big|<\overline{\kappa}_p$ for all $v\in\mathbb{R}$ and all $m=0,\ldots,M+1$. This, along with (S4.3), [24], (S2.2-3), (S4.1(i)-(iii)), (S4.2(i)) and the compactness of $\boldsymbol{C},\boldsymbol{B}$ and $\boldsymbol{A}$ implies that all the assumptions of Lemma A.1 are satisfied for $m=0,\ldots,M+1$. Therefore for all $m=0,\ldots,M+1$:

$$\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}}\Big|\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - E\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big| = O_p(1)\ \forall\ j\in\{1,\ldots,k_p\}$$

once again, using these results along with the triangle inequality we get that for all $m=0,\ldots,M+1$:

$$N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{\text{-}p}\in\boldsymbol{A}}}\Big|\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\zeta^{(m)}_{p_j}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)\Big| \le \big(N^{1-2\varepsilon}h_N^{2L}\big)^{-1/4}O_p(1) + N^{1/4}h_N^M\widetilde{D}_5$$

[23] See footnote 22.

[24] Note that the cases analyzed previously -$\widehat{S}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$, $\widehat{s}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $\widehat{s}^{(m)}_{p_N}(\pi_{-p}\mid \boldsymbol{z},\boldsymbol{\theta}_p)$- were functions of $\boldsymbol{X}_p$ only through $G_p(\cdot)$, $g_p(\cdot)$ and $g_p^{(m)}(\cdot)$ respectively, which are bounded functions everywhere in $\mathbb{R}$.

and using (S4.2(i-ii)) we have that for all $m = 0, \ldots, M+1$:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = o_p(1) \quad \forall\, j \in \{1, \ldots, k_p\} \qquad (4\star)$$

For $j \in \{1, \ldots, k_p\}$ let $\widehat{\zeta}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p}|\boldsymbol{z},\boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$. Therefore, note that we have $\widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \left(\widehat{\zeta}^{(m)}_{p_{N_1}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p), \ldots, \widehat{\zeta}^{(m)}_{p_{N_{k_p}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right)'$. By a first-order Taylor approximation we get:

$$\begin{aligned} \widehat{\zeta}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) &+ \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \left[\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right] \\ &- \frac{\widetilde{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^2} \left[\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right] \end{aligned}$$

with $\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ between $\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $\widetilde{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ between $\widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ and $f_{\boldsymbol{Z}}(\boldsymbol{z}) \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$. We know from above that by (A) and the fact that $\inf_{\boldsymbol{z}\in\boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$, we have:

$$\sup_{\boldsymbol{z}\in\boldsymbol{C}} \left| \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \right| = O_p(1)$$

By (4⋆) and the fact that $\left|\zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right| \leq \overline{X}\overline{\kappa}_p \;\forall\, (\boldsymbol{z}, \boldsymbol{\theta}_p, \pi_{-p}) \in \mathbb{S}(\boldsymbol{Z}) \times \mathbb{R}^{k_p+1} \times \mathbb{R}$, all $j \in \{1, \ldots, k_p\}$ and all $m = 0, \ldots, M+1$, we also have that for all $m = 0, \ldots, M+1$:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widetilde{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \quad \text{for all } j \in \{1, \ldots, k_p\}$$

Therefore for all $m = 0, \ldots, M+1$:

$$\begin{aligned} N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} & \left| \widehat{\zeta}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| \\ & \leq O_p(1) \cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{T}^{(m)}_{p_{N_j}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \zeta^{(m)}_{p_j}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| \\ & + O_p(1) \cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| \\ & = o_p(1) \quad \text{for all } j \in \{1, \ldots, k_p\} \qquad (\dagger) \end{aligned}$$

with the last equality following from (A) and (4⋆). Now recall that by definition we have

$$\widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \big(\widehat{\zeta}^{(m)}_{p_{N_1}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p),\dots,\widehat{\zeta}^{(m)}_{p_{N_{k_p}}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\big)'$$

$$\zeta^{(m)}_{p}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \big(\zeta^{(m)}_{p_1}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p),\dots,\zeta^{(m)}_{p_{k_p}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\big)'$$

Therefore (†) immediately implies that:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}} \left\|\widehat{\zeta}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \zeta^{(m)}_{p}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right\| = o_p(1) \qquad \forall\ m = 0,\dots,M+1$$

which proves part (E) of the lemma.

Now let $\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ be the $[j,\ell]^{th}$ elements of $\widehat{t}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $\xi_p(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ respectively, where $j,\ell \in \{1,\dots,k_p\}$. Then we have:

$$\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} X_{p_{j_n}}\boldsymbol{X}_{p_{\ell_n}}\ g^{(m)}_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z})$$

$$\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = E\big[X_{p_j}X_{p_\ell}g^{(m)}_p\big(\boldsymbol{X}'_{p}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big) \mid \boldsymbol{Z} = \boldsymbol{z}\big]$$

Take any $j,\ell \in \{1,\dots,k_p\}$, then once again using (4) and the same arguments used in the previous cases, we can show that for all $(\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A$ we have:

$$\left|E\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| = \frac{h_N^M}{M!}\left|\int\int\sum_{Q_M} u_j u_k g^{(m)}_p(\boldsymbol{u}'\boldsymbol{\beta}_p + \alpha_p\pi_{-p})(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Gamma^p_M(\boldsymbol{u},\boldsymbol{z}+h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u}\right|$$

Recall that by (S1.3), we have $|g^{(m)}_p(v)| < \overline{\kappa}_p$ for all $v \in \mathbb{R}$ and all $m = 0,\dots,M+1$, where $\overline{\kappa}_p$ is defined above. Also recall that by (S2.3) there exists $\overline{X}$ such that $\left|X_{p_j}X_{p_\ell}\right| < \overline{X}^2$ for all $j,\ell \in \{1,\dots k_p\}$ w.p.1. Therefore, for all $m = 0,\dots,M+1$ we have:

$$\left|E\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)\right| \leq \widetilde{D}_6 h_N^M \quad \forall\ (\boldsymbol{z},\boldsymbol{\theta}_p,\pi_{-p}) \in \boldsymbol{C}\times\boldsymbol{B}\times A \quad \forall\ j,\ell \in \{1,\dots,k_p\}$$

with $\widetilde{D}_6 = \frac{1}{M!}\overline{C}_2\left|Q_M\right|\overline{\kappa}_p\overline{X}^2\int\left\|\boldsymbol{\Psi}\right\|^M d\boldsymbol{\Psi}$ where $\overline{C}_2$ and $\left|Q_M\right|$ are as defined above. As we have done before, define $\boldsymbol{t} = (\boldsymbol{\theta}_p,\pi_{-p})$ and now let $\eta(X_p,\cdot,\boldsymbol{t}) = X_{p_j}X_{p_\ell}g^{(m)}_p\big(\boldsymbol{X}'_p\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)$. Take ε as described in (S4.2(i)). Recall that $\left|X_{p_j}X_{p_\ell}\right| < \overline{X}^2$ for all $j,\ell \in \{1,\dots k_p\}$ w.p.1. This, along with (S2.2-3), (S4.1(i)-(iii)), (S4.2(i)), (S4.3) and the compactness of $\boldsymbol{C}, \boldsymbol{B}$ and $\boldsymbol{A}$ implies that all the assumptions of Lemma A.1 are satisfied for $m = 0, \ldots, M+1$. Therefore for all $m = 0, \ldots, M+1$ we have:

$$\left(N^{1-\varepsilon} h_N^L\right)^{1/2} \sup_{\substack{\boldsymbol{z} \in \boldsymbol{C} \\ \boldsymbol{\theta}_p \in \boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - E\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \qquad \forall\ j, \ell \in \{1, \ldots, k_p\}$$

once again, using these results along with the triangle inequality we get:

$$N^{1/4} \sup_{\substack{\boldsymbol{z} \in \boldsymbol{C} \\ \boldsymbol{\theta}_p \in \boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right|$$
$$\leq \left(N^{1-2\varepsilon} h_N^{2L}\right)^{-1/4} O_p(1) + N^{1/4} h_N^M \widetilde{D}_5$$

and using (S4.2(i-ii)) we have that for all $m = 0, \ldots, M+1$ we have:

$$N^{1/4} \sup_{\substack{\boldsymbol{z} \in \boldsymbol{C} \\ \boldsymbol{\theta}_p \in \boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = o_p(1) \quad \forall\ j, \ell \in \{1, \ldots, k_p\} \tag{5$\star$}$$

For $j, \ell \in \{1, \ldots, k_p\}$ let $\widehat{\xi}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \frac{\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})}$. Note that $\widehat{\xi}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ is the $[j,\ell]^{th}$ element of $\widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$. By a first-order Taylor approximation we get:

$$\widehat{\xi}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) = \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) + \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \left[\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right]$$
$$- \frac{\widetilde{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^2} \left[\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right]$$

with $\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ between $\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})$ and $\widetilde{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ between $\widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ and $f_{\boldsymbol{Z}}(\boldsymbol{z}) \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$. We know from above that by (A) and the fact that $\inf_{\boldsymbol{z} \in \boldsymbol{C}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b > 0$, we have:

$$\sup_{\boldsymbol{z} \in \boldsymbol{C}} \left| \frac{1}{\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \right| = O_p(1)$$

By (5$\star$) and the fact that $\left|\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)\right| \leq \overline{X}^2 \overline{\kappa}_p\ \forall\ (\boldsymbol{z}, \boldsymbol{\theta}_p, \pi_{-p}) \in \mathbb{S}(\boldsymbol{Z}) \times \mathbb{R}^{k_p+1} \times \mathbb{R}$, all $m = 0, \ldots, M+1$ and all $j, \ell \in \{1, \ldots, k_p\}$, we also have

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{C} \\ \boldsymbol{\theta}_p \in \boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{z} \in \boldsymbol{C} \\ \boldsymbol{\theta}_p \in \boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widetilde{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| = O_p(1) \quad \text{for all } j, \ell \in \{1, \ldots, k_p\}$$

Therefore for all $m = 0, \ldots, M+1$:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{\xi}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| \leq$$

$$O_p(1) \cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{t}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right| + O_p(1) \cdot N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right|$$

$$= o_p(1) \quad \text{for all } j, \ell \in \{1, \ldots, k_p\} \qquad (\ddagger)$$

with the last equality following from (A) and (5$\star$).

Now recall that by definition, $\widehat{\xi}^{(m)}_{p_{N_{[j,\ell]}}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ and $\xi^{(m)}_{p_{[j,\ell]}}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ are the $[j,\ell]^{th}$ elements of $\widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ and $\xi^{(m)}_{p}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p)$ respectively. Therefore (‡) immediately implies that:

$$N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{C} \\ \boldsymbol{\theta}_p\in\boldsymbol{B} \\ \pi_{\text{-}p} \in \boldsymbol{A}}} \left\| \widehat{\xi}^{(m)}_{p_N}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) - \xi^{(m)}_{p}(\pi_{-p} \mid \boldsymbol{z}, \boldsymbol{\theta}_p) \right\| = o_p(1) \qquad \forall\ m = 0, \ldots, M+1$$

which proves part (E) of the lemma and completes its proof. $\square$

Take $\boldsymbol{z} \in \mathbb{S}(\boldsymbol{Z})$, $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$ and $(\pi_1, \pi_2) \in \mathbb{R}^2$. From here on, we will denote:

$$\boldsymbol{\pi} \equiv (\pi_1, \pi_2) \in \mathbb{R}^2$$

$$\boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv \big(\varphi_1(\boldsymbol{\pi}_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1), \varphi_2(\boldsymbol{\pi}_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2)\big)' \in \mathbb{R}^2$$

$$\boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv -\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)'\big(\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big) \in \mathbb{R}$$

$$\widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv \big(\widehat{\varphi}_{1_N}(\boldsymbol{\pi}_2 \mid \boldsymbol{z}, \boldsymbol{\theta}_1), \widehat{\varphi}_{2_N}(\boldsymbol{\pi}_1 \mid \boldsymbol{z}, \boldsymbol{\theta}_2)\big)' \in \mathbb{R}^2$$

$$\widehat{\boldsymbol{Q}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) \equiv -\big(\boldsymbol{\pi} - \widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big)'\big(\boldsymbol{\pi} - \widehat{\boldsymbol{\varphi}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})\big) \in \mathbb{R}$$

For $(\boldsymbol{z}, \boldsymbol{\theta}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ let $\big(\pi_1^*(\boldsymbol{z}, \boldsymbol{\theta}), \pi_2^*(\boldsymbol{z}, \boldsymbol{\theta})\big)' \equiv \boldsymbol{\pi}^*(\boldsymbol{z}, \boldsymbol{\theta})$ denote the solution (for π_1 and π_2) to the system

$$\boldsymbol{\pi} - \boldsymbol{\varphi}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0}$$

Then, by (S3.2) and Theorem 4.1 we know that for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ there exists a unique such $\boldsymbol{\pi}^*(\boldsymbol{z}, \boldsymbol{\theta})$. By (S3.2), we also have that:

$$\forall (\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}:\ \boldsymbol{\pi}^* - \boldsymbol{\varphi}(\boldsymbol{\pi}^* \mid \boldsymbol{z}, \boldsymbol{\theta}) = \mathbf{0} \quad \text{if and only if} \quad \boldsymbol{\pi}^* = \operatorname*{argmax}_{\boldsymbol{\pi}\in\mathbb{R}^2} \boldsymbol{Q}(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$$

By (S1.3) and (S2.3) (see Theorem 4.1) we also know that $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ is strictly inside $[0,1]^2$ for all $\boldsymbol{z} \in \mathbb{S}(\boldsymbol{Z})$ and all $\boldsymbol{\theta} \in \mathbb{R}^{k+2}$. In particular, since $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\mathcal{Z}}$ is a compact set in the interior of $\mathbb{S}(\boldsymbol{Z})$, there exists a $0 < \tau < 1$ such that $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in [\tau, 1-\tau]^2$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$

The next result establishes uniform consistency of the proposed estimator $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta})$ in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Lemma A.3 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and let $\big(\widehat{\pi^*_1}_N(\boldsymbol{z}, \boldsymbol{\theta}), \widehat{\pi^*_2}_N(\boldsymbol{z}, \boldsymbol{\theta})\big)' \equiv \widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta})$ satisfy:*

$$\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta}) = \underset{\boldsymbol{\pi} \in [0,1]^2}{argmax}\, \widehat{\boldsymbol{Q}}_N(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta})$$

Then

$$\underset{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}}{sup} \Big\| \widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta}) - \boldsymbol{\pi}^*(\boldsymbol{z}, \boldsymbol{\theta}) \Big\| = o_p(1)$$

Proof:

Take $\delta > 0$ and for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ let $M_{\boldsymbol{\theta},\boldsymbol{z}} = \big\{\boldsymbol{\pi} : \big\|\boldsymbol{\pi} - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big\| < \delta\big\}$ and let $\overline{M}_{\boldsymbol{\theta},\boldsymbol{z}}$ be the complement of $M_{\boldsymbol{\theta},\boldsymbol{z}}$ in $\mathbb{R}^2$. Now define the set $\mathcal{N}_{\boldsymbol{\theta},\boldsymbol{z}} = \overline{M}_{\boldsymbol{\theta},\boldsymbol{z}} \cap [0,1]^2$. Then $\mathcal{N}_{\boldsymbol{\theta},\boldsymbol{z}} \in [0,1]^2$ is compact for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, and by continuity we get that $\max_{\boldsymbol{\pi} \in \mathcal{N}_{\boldsymbol{\theta},\boldsymbol{z}}} Q(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z})$ exists for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Now define $\varepsilon = \inf_{\substack{\boldsymbol{\theta} \in \boldsymbol{\Theta} \\ \boldsymbol{z} \in \mathcal{Z}}} \big[Q(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{\theta}, \boldsymbol{z}) - \max_{\boldsymbol{\pi} \in \mathcal{N}_{\boldsymbol{\theta},\boldsymbol{z}}} Q(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z})\big]$. Then $\varepsilon > 0$, since for each $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{Z}$ we have that $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$ is the unique solution to $\max_{\boldsymbol{\pi} \in \mathbb{R}^2} Q(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z})$ (see Theorem 4.1). Now let A_N be the event:

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{Z} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta} \\ \boldsymbol{\pi} \in [0,1]^2}} \Big| \widehat{Q}_N(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z}) - Q(\boldsymbol{\pi} \mid \boldsymbol{\theta}, \boldsymbol{z}) \Big| < \frac{\varepsilon}{2}$$

we know $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. By definition of $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta})$, we also have $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z}, \boldsymbol{\theta}) \in [0,1]^2$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, we have the following implications:

$$A_N \Rightarrow Q\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{\theta}, \boldsymbol{z}\big) > \widehat{Q}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{\theta}, \boldsymbol{z}\big) - \frac{\varepsilon}{2} \qquad \forall\, (\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$$

$$A_N \Rightarrow \widehat{Q}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{\theta}, \boldsymbol{z}\big) > Q\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{\theta}, \boldsymbol{z}\big) - \frac{\varepsilon}{2} \qquad \forall\, (\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$$

By definition of $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{z},\boldsymbol{\theta})$ we also have $\widehat{Q}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{\theta},\boldsymbol{z}\big) \geq \widehat{Q}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{\theta},\boldsymbol{z}\big)\ \forall\ (\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Combining this with the two implications outlined above, we get:

$$A_N \Rightarrow Q\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{\theta},\boldsymbol{z}\big) > Q\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{\theta},\boldsymbol{z}\big) - \varepsilon \quad \forall\ (\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$$

By definition of ε, we can conclude that $A_N \Rightarrow \big\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big\| < \delta$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ or equivalently: $A_N \Rightarrow \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\big\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big\| < \delta$. As a consequence, we then have that $\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\big\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big\| < \delta\right\} \geq \Pr(A_N)$. Now, by Lemma A.2(B) we know that $\Pr(A_N) \to 1$. Therefore $\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\big\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big\| < \delta\right\} \to 1$. Since δ is an arbitrary positive number, this implies that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big\| = o_P(1)$, as claimed. □

Let $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ and $J(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ denote the Jacobian with respect to $\boldsymbol{\pi}$ of $\boldsymbol{\pi} - \widehat{\varphi}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ and $\boldsymbol{\pi} - \varphi(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ respectively. Then $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ is given by:

$$\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) = \begin{pmatrix} 1 & -\alpha_1\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}) \\ -\alpha_2\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}) & 1 \end{pmatrix}$$

while $J(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ is given by:

$$J(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) = \begin{pmatrix} 1 & -\alpha_1\delta_1(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}) \\ -\alpha_2\delta_2(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}) & 1 \end{pmatrix}$$

We will let $\widehat{d}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ and $d(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ denote the determinants of $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ and $J(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$ respectively. Therefore, we have $\widehat{d}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) = 1 - \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2)$ and $d(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) = 1 - \alpha_1\alpha_2\delta_1(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\delta_2(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2)$. The next lemma establishes uniform convergence in probability of $\widehat{J}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})^{-1}$ in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Assumption (S3.2) -which also guarantees uniqueness of equilibrium- plays a crucial role for this result.

Lemma A.4 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Then*

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left\| \widehat{J}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1} - J(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1}\right\| = o_p(N^{-1/4})$$

Proof: We begin by showing that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})\right|^{-1} = O_p(1)$. To see this, note first that by (S3.2), there exists $0<\underline{d}<\infty$ such that $\alpha_1\alpha_2\delta_1(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\delta_2(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}_2)<1-\underline{d}$ for all $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and consequently $d(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})>\underline{d}$ for all $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Now, by Lemma A.2(C), we have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{\delta}_{1_N}(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}_2) - \delta_1(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\delta_2(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}_2)\right| = o_p(\mathrm{N}^{-1/4}) \qquad (\triangle 1)$$

and therefore $\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}_2)<1-\underline{d}\right\}\to 1$, whence we obtain:

$$\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})\right|^{-1} > \underline{d}^{-1}\right\} = \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})\right| < \underline{d}\right\}$$

$$= \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}_2) > 1-\underline{d}\right\}\to 0$$

and consequently: $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\ \widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})\ \right|^{-1} = O_p(1)$.

From $(\triangle 1)$ we have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta}) - d(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})\right| = o_p(\mathrm{N}^{-1/4})$. Using (S3.2), this yields:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1} - d(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1}\right| = o_p(\mathrm{N}^{-1/4}) \qquad (\triangle 2)$$

Combining ($\triangle 2$) with Lemma A.2 (C) we also have:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left|\widehat{d}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1}\widehat{\delta}_{p_N}(\pi_{-p}\mid\boldsymbol{z},\boldsymbol{\theta}_p) - d(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1}\delta_p(\pi_{-p}\mid\boldsymbol{z},\boldsymbol{\theta}_p)\right| = o_p(\mathrm{N}^{-1/4}) \quad \text{for } p\in\{1,2\}$$

which combined with ($\triangle 2$) implies that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \left\| \widehat{J}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1} - J(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})^{-1}\right\| = o_p(\mathrm{N}^{-1/4})$ as claimed. □

We next use the previous lemmas to establish a precise rate of uniform convergence of $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$.

Lemma A.5 (Lemma 4.2(A)) *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and let $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$be as defined in Lemma A.3. Then*

$$\mathop{sup}_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(N^{-1/4})$$

Proof: The steps resemble those of the proof of Theorem 3.1 in Newey and McFadden (1994). First, take $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and define the indicator variables:

$$\widehat{1}_N(\boldsymbol{\theta},\boldsymbol{z}) = \mathbb{1}\left\{\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\in(0,1)^2\right\}$$
$$\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) = \mathbb{1}\left\{\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\in(0,1)^2 \quad \text{and} \quad \widehat{d}_N\left(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\right)\neq 0\right\}$$

Notice that because by definition we have $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\in[0,1]^2$, then $\widehat{1}_N(\boldsymbol{\theta},\boldsymbol{z})=0$ if and only if $\widehat{\boldsymbol{\pi}^*_{p_N}}(\boldsymbol{\theta},\boldsymbol{z})$ equals zero or one for some $p\in\{1,2\}$. If $\widehat{1}_N(\boldsymbol{\theta},\boldsymbol{z})=1$, then $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ satisfies the first order conditions

$$\widehat{J}_N\left(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\right)\left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \widehat{\varphi}_N\left(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\right)\right] = \mathbf{0}$$

Now, if $\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z})=1$ then $\widehat{J}_N\left(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\right)$ is invertible and $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ satisfies the first order conditions if and only if $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \widehat{\varphi}_N\left(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\right)=\mathbf{0}$. Therefore, $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ is defined by the first-order conditions:

$$\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z})\Big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \widehat{\varphi}_N\Big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big)\Big] = \mathbf{0} \tag{1$\diamond$}$$

By a mean-value expansion theorem, we have:

$$\overline{1}\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}),\boldsymbol{\theta}\big)\widehat{J}_N\big(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big] = \overline{1}\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}),\boldsymbol{\theta}\big)\Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big]$$

where $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ is equal to the mean value (between $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ and $\boldsymbol{\pi}^*_N(\boldsymbol{\theta},\boldsymbol{z})$) if $\overline{1}(\cdot) = 1$ and is equal to $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})$ otherwise. Now define one more indicator variable:

$$\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z}) = \mathbb{1}\Big\{\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \in (0,1)^2\ , \quad \widehat{d}_N\Big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big) \neq 0 \quad \text{and} \quad \widehat{d}_N\Big(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big) \neq 0\Big\}$$

Then, the mean-value approximation becomes:

$$\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})\Big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big] = \widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})\widehat{J}_N\Big(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big)^{-1}\Big[\widehat{\varphi}_N\Big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big]$$

and we get:

$$\begin{aligned} N^{1/4}\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] = {} & \widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})\widehat{J}_N\Big(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\Big)^{-1}N^{1/4}\big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] \\ & + N^{1/4}\big[1 - \widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})\big]\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] \end{aligned}$$

By definition, $\left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\right|$ can only equal zero or one. In fact, $\sup\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\right| = 1$ only if any of the following holds:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \big(\widehat{\pi^*_{N_p}}(\boldsymbol{\theta},\boldsymbol{z})\big) \geq 1 \quad \text{or} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \big(-\widehat{\pi^*_{N_p}}(\boldsymbol{\theta},\boldsymbol{z})\big) \geq 0 \quad \text{for some } p \in \{1,2\}$$

$$\text{or} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta} \\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2) \geq 1$$

where the last condition follows from the fact that $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ and $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. This implies that:

$$\begin{aligned}
\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})-1\right| = 1\right\} \leq{} & \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(\widehat{\pi^*_{N_1}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 1\right\} + \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(-\widehat{\pi^*_{N_1}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 0\right\} \\
&+ \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(\widehat{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 1\right\} + \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(-\widehat{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 0\right\} \\
&+ \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2) \geq 1\right\}
\end{aligned}$$

By (S1.3) and (S2.3) (see Lemma 4.1) we know that since $\boldsymbol{\Theta}$ is compact and $\boldsymbol{\mathcal{Z}}$ is a compact set in the interior of $\mathbb{S}(\boldsymbol{Z})$, there exists a $0 < \tau < 1$ such that $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \in [\tau, 1-\tau]^2$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. This implies that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right) = 1-\tau < 1$ and $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right) = -\tau < 0$. By Lemma A.3 we know that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z},\boldsymbol{\theta}) - \boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta})\right\| = o_p(1)$, these results together imply that:

$$\begin{aligned}
&\left[\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(\widehat{\pi^*_{N_1}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 1\right\} + \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(-\widehat{\pi^*_{N_1}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 0\right\} + \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(\widehat{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 1\right\}\right. \\
&\qquad\left. + \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left(-\widehat{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z})\right) \geq 0\right\}\right] \to 0
\end{aligned}$$

Also, using the proof of Lemma A.4, we have: [25]

$$\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2) \geq 1\right\} \to 0$$

and therefore

$$\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z})-1\right| = 1\right\} \to 0$$

---

[25] There we showed that $\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}\\ \boldsymbol{\pi}\in[0,1]^2}} \alpha_1\alpha_2\widehat{\delta}_{1_N}(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\widehat{\delta}_{2_N}(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2) < 1-\underline{d}\right\} \to 1$, where $0 < \underline{d} < \infty$ is such that $\alpha_1\alpha_2\delta_1(\pi_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\delta_2(\pi_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2) < 1-\underline{d}$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. The existence of such $\underline{d}$ is guaranteed by (S3.2).

Now let $B_N = N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\right|$ and denote $p_N \equiv \Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widetilde{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\right| = 1\right\}$. Then we have:

$$B_N = \begin{cases} N^{1/4} & \text{with probability } \ p_N \\ 0 & \text{with probability } \ 1-p_N \end{cases}$$

Now take any $M > 0$. Then:

$$\Pr\big[B_N \geq M\big] = \begin{cases} 0 & \text{if } N < M^4 \\ p_N & \text{if } N \geq M^4 \end{cases}$$

therefore, since $p_N \to 0$ we have $B_N = O_p(1)$ [26]. By Lemma A.4 and the fact that $\widetilde{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$, we also have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{J}_N\big(\widetilde{\pi^*_{N_2}}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\right\| = O_p(1)$. Using these results, we have:

$$\begin{aligned} N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \leq O_p(1)\ & N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \\ & + O_p(1) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \end{aligned}$$

now, we have:

$$\begin{aligned} N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \leq \quad & N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\right\| \\ & + N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \\ & = o_p(1) + 0 \end{aligned}$$

where the last equality comes from Lemma A.2(B) and the fact that $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and also from the fact that $\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = \mathbf{0}$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ by the game's equilibrium conditions. Now, by Lemma A.3, we have

[26] In fact, the argument shows the stronger result that $B_N = o_p(1)$.

$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(1)$. These results together imply that:

$$N^{1/4}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| \leq o_p(1)$$

and therefore $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\left\|\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| = o_p(\mathrm{N}^{-1/4})$ as claimed. □

Lemma A.6 *Let* $\boldsymbol{\mathcal{Z}}$ *be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take* $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ *and let* $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$*be as defined in Lemma A.3. Then:*

(A) $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big]+o_p(N^{-1/2})$

for all $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$*.*

(B) As in the proof of Lemma A.2(B), define:

$$\widehat{S}_{p_N}(\pi_{-p}\mid\boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N}G_p\big(\boldsymbol{X}'_{p_n}\boldsymbol{\beta}_p+\alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n-\boldsymbol{z}) \quad \textit{for } p\in\{1,2\}$$

and let $\widehat{S}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta}) = \big(\,\widehat{S}_{1_N}(\boldsymbol{\pi}_2\mid\boldsymbol{z},\boldsymbol{\theta}_1)\,,\,\widehat{S}_{2_N}(\boldsymbol{\pi}_1\mid\boldsymbol{z},\boldsymbol{\theta}_2)\,\big)'$*. Then:*

$$\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)-\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big]+o_p(N^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z})\in\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$*.*

Proof: We will use the same notation as in Lemma A.2. First define:

$$\underset{2\times4}{\widehat{H}_N(\boldsymbol{\pi}\mid\boldsymbol{z},\boldsymbol{\theta})} = \begin{pmatrix} 0 & 0 & 0 & \alpha_1^2\widehat{\zeta}^{(0)}_{1_N}(\pi_2\mid\boldsymbol{z},\boldsymbol{\theta}) \\ \alpha_2^2\widehat{\zeta}^{(0)}_{2_N}(\pi_1\mid\boldsymbol{z},\boldsymbol{\theta}) & 0 & 0 & 0 \end{pmatrix}$$

From (1⋄) we have $\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z})\Big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)\Big]=\boldsymbol{0}$. A second-order approximation yields -after rearranging-:

$$\begin{aligned}\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z})\widehat{J}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] &= \overline{1}_N(\boldsymbol{\theta},\boldsymbol{z})\Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \\ &+\frac{1}{2}\,\widehat{H}_N(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta})\big\{\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big]\otimes\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big]\big\}\Big]\end{aligned}$$

where $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ is between $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})$ if $\overline{1}_N(\cdot)=1$ and is equal to $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})$ otherwise. By assumption (S2.3), Lemma A.2(E) and the fact that $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\in[0,1]^2$ (a compact set) for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, we have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \widehat{H}_N(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}) \right\| = O_p(1)$. Using Theorem A.5 we also have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \otimes \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \right\| = o_p(\mathrm{N}^{-1/2})$ and therefore:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \widehat{H}_N(\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}) \left\{ \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \otimes \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \right\} \right\| = o_p(\mathrm{N}^{-1/2})$$

Consequently, we get:

$$\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \widehat{J}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] = \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \left[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Adding and subtracting $\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right]$, we get:

$$\begin{aligned} \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] &= \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \left[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \\ &+ \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \left[J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \widehat{J}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\right] \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

From Lemma A.2 (C) and Theorem A.5 we have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \left[J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \widehat{J}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\right] \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \right\| = o_p(\mathrm{N}^{-1/2})$$

and therefore:

$$\begin{aligned} J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] &= \left[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] \\ &+ \left[\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) - 1\right] \left[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] + \left[1 - \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z})\right] \left[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\right] + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Let $C_N = N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left| \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) - 1 \right|$. By the same arguments as those of the proof of Theorem A.5, we have $C_N = O_p(1)$. We also showed there (using Lemma A.2 and the equilibrium conditions) that $N^{1/4} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| = o_p(1)$. Using these facts along with the main result of Theorem A.5, we have:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\{ \left| \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) - 1 \right| \left\| \widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| \right\} = o_p(\mathrm{N}^{-1/2})$$

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\{ \left| 1 - \overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) \right| \left\| \widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \right\| \right\} = o_p(\mathrm{N}^{-1/2})$$

which yields:

$$J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\big[\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] \;=\; \big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big] + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. By $(S3.2)$, we have that $J\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}$ exists and satisfies the condition $\Big\|J\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\| < \overline{D}$ for some $\overline{D} < \infty$ for all $(\boldsymbol{\pi},\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2\times\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Since $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ for all $(\boldsymbol{\theta},\boldsymbol{z})\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$, we then have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\| < \overline{D}$. Consequently, we have:

$$\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \;=\; J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big] + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$, which proves part (A) of the result.

To prove part (B) note first that, by definition: $\widehat{\varphi}_{N_p}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \widehat{S}_{N_p}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p)/\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ for $p \in \{1,2\}$.

Performing second-order approximation of $\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)/\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ around

$$\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) = f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \quad \text{and} \quad \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) = f_{\boldsymbol{Z}}(\boldsymbol{z})$$

yields:

$$\begin{aligned}
\widehat{\varphi}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \;&=\; \varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\\
&+ \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\begin{pmatrix}1 & -\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\end{pmatrix} \times \begin{pmatrix}\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\\ \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\end{pmatrix}\\
&+ \begin{pmatrix}\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\\ \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\end{pmatrix}' \widetilde{\Upsilon_{Np}}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\\
&\times \begin{pmatrix}\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\\ \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\end{pmatrix}
\end{aligned}$$

where $\widetilde{\Upsilon_{Np}}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \begin{pmatrix}0 & -\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^{-2}\\ -\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^{-2} & 2\widetilde{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})^{-3}\end{pmatrix}$

with $\widetilde{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ between $\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p)$ and $\widetilde{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ between $\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})$ and $f_{\boldsymbol{Z}}(\boldsymbol{z})$. From the proof of Lemma A.2 (B) and the fact that $\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, we know that

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \right\| = o_p(\mathrm{N}^{-1/4})$$

from Lemma A.2 (A) also know that $\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left| \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \right| = o_p(\mathrm{N}^{-1/4})$. Using the fact that there exists $b > 0$ such that $\inf_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} f_{\boldsymbol{Z}}(\boldsymbol{z}) > b$, these two results imply that

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \widetilde{\Upsilon_{N}}_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \right\| = O_p(1)$$

and consequently

$$\widehat{\varphi}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \ \varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) =$$

$$\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \begin{pmatrix} 1 & -\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \end{pmatrix} \times \begin{pmatrix} \widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - f_{\boldsymbol{Z}}(\boldsymbol{z})\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) \\ \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z}) \end{pmatrix} + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

From the equilibrium conditions we have: $\varphi_p(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \pi^*_p(\boldsymbol{\theta},\boldsymbol{z})$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, for $p \in \{1,2\}$ the second-order approximation yields:

$$\widehat{\varphi}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \pi^*_p(\boldsymbol{\theta},\boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \left[\widehat{S}_{N_p}(\pi^*_{\text{-}p}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\pi^*_p(\boldsymbol{\theta},\boldsymbol{z}) \right] + o_p(\mathrm{N}^{-1/2})$$

which immediately implies that:

$$\widehat{\varphi}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \left[\widehat{S}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \right] + o_p(\mathrm{N}^{-1/2})$$

since by definition we have: $\widehat{\varphi}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) \equiv \big(\widehat{\varphi}_{N_1}(\pi^*_2(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_1), \widehat{\varphi}_{N_2}(\pi^*_1(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_2)\big)'$, $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \equiv \big(\pi^*_1(\boldsymbol{\theta},\boldsymbol{z}), \pi^*_2(\boldsymbol{\theta},\boldsymbol{z})\big)'$, $\widehat{S}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) \equiv \big(\widehat{S}_{N_1}(\pi^*_2(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_1), \widehat{S}_{N_2}(\pi^*_1(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_2)\big)'$. and using the result from part (A) -above- we finally get:

$$\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1} \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \Big[\widehat{S}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \Big] + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. This proves part (B) and completes the proof. $\square$

The following result is a consequence of Lemmas A.2(B-F) and Theorem A.5.

Lemma A.7 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and let $\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{z},\boldsymbol{\theta}) = \left(\widehat{\pi^*_1}_N(\boldsymbol{z},\boldsymbol{\theta}), \widehat{\pi^*_2}_N(\boldsymbol{z},\boldsymbol{\theta})\right)$ be as defined in Lemma A.3. Then for $p \in \{1, 2\}$ we have:*

$$(A) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\varphi}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \varphi_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| = o_p(N^{-1/4})$$

$$(B) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \delta_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| = o_p(N^{-1/4})$$

$$(C) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}^{(m)}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \delta^{(m)}_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| = o_p(N^{-1/4}) \quad m = 1, \ldots, M$$

$$(D) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\zeta}^{(m)}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \zeta^{(m)}_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right\| = o_p(N^{-1/4}) \quad m = 0, \ldots, M$$

$$(E) \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\xi}^{(m)}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \xi^{(m)}_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right\| = o_p(N^{-1/4}) \quad m = 0, \ldots, M$$

Proof: A mean-value approximation yields:

$$\widehat{\varphi}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) = \widehat{\varphi}_{p_N}\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) + \alpha_p\ \widehat{\delta}_{p_N}\left(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\left[\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) - \pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})\right]$$

where $\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta})$ is between $\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta})$ and $\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})$. We have $\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \in [0,1]$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, using Lemma A.2 (C) we have:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}_{p_N}\left(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| = O_p(1) \quad \text{for } p \in \{1,2\}$$

Combining this with Theorem A.5 we get

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\{\left|\alpha_p\ \widehat{\delta}_{p_N}\left(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| \cdot \left|\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) - \pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})\right|\right\} = o_p(\mathrm{N}^{-1/4}) \quad \text{for } p \in \{1,2\}$$

and therefore:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\varphi}_{p_N}\left(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \varphi_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| \le \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\varphi}_{p_N}\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \varphi_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| + o_p(\mathrm{N}^{-1/4}) \quad \text{for } p \in \{1,2\}$$

We have $\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \in [0,1]$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Therefore, using Lemma A.2 (B) we have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\varphi}_{p_N}\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right) - \varphi_p\left(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\right)\right| = o_p(\mathrm{N}^{-1/4})$ for $p \in \{1,2\}$

and consequently, $\sup\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\varphi}_{p_N}\big(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \varphi_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right| = o_p(\mathrm{N}^{-1/4})$ for $p \in \{1,2\}$, which proves part (A).

To show part (B) we proceed similarly. A mean-value approximation yields:

$$\widehat{\delta}_{p_N}\big(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) = \widehat{\delta}_{p_N}\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) + \alpha_p\, \widehat{\delta}^{(1)}_{p_N}\big(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\big[\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) - \pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})\big]$$

with $\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta})$ between $\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta})$ and $\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})$. As above, we have $\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \in [0,1]$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Therefore, using Lemma A.2(D):

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}^{(1)}_{p_N}\big(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right| = O_p(1) \quad \text{for } p \in \{1,2\}$$

which combined with Theorem A.5 yields:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}^{(1)}_{p_N}\big(\widetilde{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\big[\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) - \pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta})\big]\right| = o_p(\mathrm{N}^{-1/4})$$

using the fact that $\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \in [0,1]$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ we have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}_{p_N}\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right| = o_p(\mathrm{N}^{-1/4})$$

combining these results we get: $\sup\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}_{p_N}\big(\widehat{\pi^*_{\text{-}p}}_N(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right| = o_p(\mathrm{N}^{-1/4})$, which shows part (B).

The proof of parts (C)-(E) is done following the same steps: starting from a mean-value approximation, and using Theorem A.5 along with Lemma A.2 (D)-(F), which -as was the case in the paragraphs above- are applicable because the mean values are always in the set $[0,1]$, which is compact. □

Lemma A.8 (Proof of Lemma 4.2(B)) *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ and let $\big(\widehat{\pi^*_1}_N(\boldsymbol{z},\boldsymbol{\theta}), \widehat{\pi^*_2}_N(\boldsymbol{z},\boldsymbol{\theta})\big)$*

be as defined in Lemma A.3. Then we have:

$$\text{(A)} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \right\| = o_p(N^{-1/4})$$

$$\text{(B)} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \right\| = o_p(N^{-1/4})$$

Proof: As in the proof of Theorem A.5, define the indicator function

$$\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) = \mathbb{1}\Big\{ \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \in (0,1)^2 \quad \text{and} \quad \widehat{d}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) \neq 0 \Big\}$$

If $\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) = 1$, then $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ satisfies $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) = \mathbf{0}$ (see 1$\diamond$ above). $\widehat{d}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) \neq 0$ implies that the assumptions of the Implicit Function Theorem are satisfied and we have:

$$\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) = \widehat{J}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)$$

where $\underset{2\times(k+2)}{\nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)} = \Big(\underset{2\times(k_1+1)}{\nabla_{\boldsymbol{\theta_1}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)} \,,\, \underset{2\times(k_2+1)}{\nabla_{\boldsymbol{\theta_2}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)} \Big)$, with:

$$\nabla_{\boldsymbol{\theta}_1} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) = \begin{pmatrix} \underset{1\times k_1}{\widehat{\zeta}_{1_N}^{(0)}\big(\widehat{\pi_{2_N}^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_1\big)'} \,, & \underset{1\times 1}{\widehat{\pi_{2_N}^*}(\boldsymbol{\theta},\boldsymbol{z})\ \widehat{\delta}_{1_N}\big(\widehat{\pi_{2_N}^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_1\big)} \\ \underset{1\times k_1}{\mathbf{0}'} & \underset{1\times 1}{0} \end{pmatrix}$$

$$\nabla_{\boldsymbol{\theta}_2} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) = \begin{pmatrix} \underset{1\times k_2}{\mathbf{0}'} & \underset{1\times 1}{0} \\ \underset{1\times k_2}{\widehat{\zeta}_{2_N}^{(0)}\big(\widehat{\pi_{1_N}^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_2\big)'} \,, & \underset{1\times 1}{\widehat{\pi_{1_N}^*}(\boldsymbol{\theta},\boldsymbol{z})\ \widehat{\delta}_{2_N}\big(\widehat{\pi_{1_N}^*}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_2\big)} \end{pmatrix}$$

In the proof of Theorem A.5 we also showed that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \big|\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\big| = o_p(1)$ [27], which is a consequence of Lemma A.3. This implies that

$$\Pr\Big\{ \overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) \neq 1 \quad \text{for some } (\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}} \Big\} \leq \Pr\Big\{ \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) \neq 1 \Big\} \to 0$$

[27] In fact, we showed that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} N^{1/4}\big|\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\big| = o_p(1)$ (see footnote 26). Using the same arguments, we can extend this result and show that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} f(N)\big|\overline{1}_N(\boldsymbol{\theta},\boldsymbol{z}) - 1\big| = o_p(1)$ for any increasing function $f(\cdot)$.

therefore, with probability approaching one uniformly in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, we have

$$\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) = \widehat{J}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$$

therefore, using Theorem A.5 and Lemmas A.7(B-C), we have:

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big\| \nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \Big\| = o_p(\mathrm{N}^{-1/4})$$

following the notation used above (see Lemma A.4) denote

$$\widehat{d}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) = 1 - \alpha_1 \alpha_2 \widehat{\delta}_{1_N}\big(\widehat{\pi^*_{2_N}}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_1\big) \widehat{\delta}_{2_N}\big(\widehat{\pi^*_{1_N}}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_2\big)$$

then, a mean-value approximation along with assumption (S3.2) and Lemma A.7 (B) yields:

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big| \widehat{d}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} - d\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \Big| = o_p(\mathrm{N}^{-1/4})$$

note that

$$\widehat{J}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} = \widehat{d}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \begin{pmatrix} 1 & -\alpha_1 \widehat{\delta}_{1_N}\big(\widehat{\pi^*_{N_2}}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_1\big) \\ -\alpha_2 \widehat{\delta}_{2_N}\big(\widehat{\pi^*_{N_1}}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_2\big) & 1 \end{pmatrix}$$

and therefore, using the above result with Lemma A.7 (B) we get

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big\| \widehat{J}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \Big\| = o_p(\mathrm{N}^{-1/4})$$

and consequently,

$$\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big\| \widehat{J}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \Big\| = o_p(\mathrm{N}^{-1/4})$$

From the equilibrium conditions and assumption (S3.2), the Implicit Function Theorem holds for the equilibrium conditions $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) - \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and consequently [28]: $\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$ for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and therefore we have $\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big\| \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \Big\| = o_p(\mathrm{N}^{-1/4})$, which proves part (A) of

[28]See Lemma 4.1.

the Lemma. To show part (B), note first that in the proof of Lemma 4.1 we showed that: $\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})} = \nabla_{\boldsymbol{\theta}}\Big(\text{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\}\Big)$ for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. By the argument outlined above, with probability approaching one uniformly in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$ we have:

$$\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta\theta}'}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})} = \nabla_{\boldsymbol{\theta}}\Big(\text{vec}\,\Big\{\widehat{J}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\}\Big)$$

which depends on the terms:

$$\widehat{d}_N\big(\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}, \quad \widehat{\delta}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta})\widehat{\zeta}^{(0)}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$
$$\widehat{\xi}^{(1)}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \widehat{\zeta}^{(1)}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta})\widehat{\zeta}^{(1)}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$
$$\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta})^2\widehat{\delta}^{(1)}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \nabla_{\boldsymbol{\theta}}\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta})\widehat{\delta}_{p_N}\big(\widehat{\pi^*_{\text{-}p_N}}(\boldsymbol{z},\boldsymbol{\theta}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

for $p = 1, 2$. Therefore, using Lemma A.7 (C)-(E) and Theorem A.5 along with part (A) of the present Lemma -shown above- we get $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\nabla_{\boldsymbol{\theta\theta}'}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big\| = o_p(\text{N}^{-1/4})$. □

Lemma A.9 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Then there exist matrices* $\underset{2\times(k+2)}{\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)}$ *and* $\underset{2\times(k+2)}{\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)}$ *such that*

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - f_{\boldsymbol{Z}}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\| = o_p(N^{-1/4})$$

and

$$\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big] + o_p(N^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$.

Proof:

Let $\underset{(2(k+2)\times 2)}{\nabla_{\boldsymbol{\theta\pi}'}\varphi(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})} = \nabla_{\boldsymbol{\pi}}\text{vec}\Big(\nabla_{\boldsymbol{\theta}}\varphi(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})\Big)$ and $\underset{(2(k+2)\times 2)}{\nabla_{\boldsymbol{\theta\pi}'}\widehat{\varphi}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})} = \nabla_{\boldsymbol{\pi}}\text{vec}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})\Big)$ from Lemma A.2 and the fact that $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \in [0,1]^2$ (a compact set) for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$, we have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\nabla_{\boldsymbol{\theta\pi}'}\widehat{\varphi}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta\pi}'}\varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})\Big\| = o_p(\text{N}^{-1/4})$. Using Lemma A.2 and

Theorem A.5 and A.7 we can also show that [29] $\sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}} \Big\| \nabla_{\theta\pi'}\widehat{\varphi}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta) - \nabla_{\theta\pi'}\varphi(\pi^*(\theta,z) \mid z,\theta)\Big\| = o_p(\mathrm{N}^{-1/4})$ Let $\mathbf{I}_{k+2}$ denote an $(k+2)\times(k+2)$ identity matrix. Therefore, a second-order approximation, along with Lemmas A.2, A.7 and A.5 yields:

$$\nabla_{\theta}\widehat{\varphi}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta) - \nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta) = \nabla_{\theta}\widehat{\varphi}_N(\pi^*(\theta,z) \mid z,\theta) - \nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta)$$
$$+ \nabla_{\theta\pi'}\varphi(\pi \mid z,\theta)'\Big\{\mathbf{I}_{k+2} \otimes \big[\widehat{\pi^*_N}(\theta,z) - \pi^*(\theta,z)\big]\Big\} + o_p(\mathrm{N}^{-1/2})$$

for all $(\theta,z) \in \Theta \times \mathcal{Z}$. Let

$$\underset{(1\times(k_p+1))}{\widehat{A}_{N_{(p)}}}(\pi_{\text{-}p} \mid z,\theta_p) = \Big(\underset{(1\times k_p)}{\widehat{T}^{(0)}_{p_N}(\pi_{\text{-}p} \mid z,\theta_p)'}, \quad \pi_{\text{-}p}\ \underset{(1\times 1)}{\widehat{s}_{p_N}(\pi_{\text{-}p} \mid z,\theta_p)}\Big)$$

$$\underset{((k+2)\times(k+2))}{\widehat{A}_N(\pi \mid z,\theta)} = \begin{pmatrix} \underset{(1\times(k_1+1))}{\widehat{A}_{N_{(1)}}(\pi_2 \mid z,\theta_1)}, & \underset{(1\times(k_2+1))}{\mathbf{0}} \\ \underset{(1\times(k_1+1))}{\mathbf{0}} & \underset{(1\times(k_2+1))}{\widehat{A}_{N_{(2)}}(\pi_1 \mid z,\theta_2)} \end{pmatrix}$$

Then, using a second-order approximation we get:

$$\nabla_{\theta}\widehat{\varphi}_N(\pi^*(\theta,z) \mid z,\theta) - \nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta) =$$
$$\frac{1}{f_Z(z)}\Big[\widehat{A}_N(\pi^*(\theta,z) \mid z,\theta) - \widehat{f}_{Z_N}(z)\nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta)\Big] + o_p(\mathrm{N}^{-1/2})$$

for all $(\theta,z) \in \Theta \times \mathcal{Z}$. Therefore, using Lemma A.6(B) we obtain:

$$\nabla_{\theta}\widehat{\varphi}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta) - \nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta) = \frac{1}{f_Z(z)}\Big[\widehat{A}_N(\pi^*(\theta,z) \mid z,\theta) - \widehat{f}_{Z_N}(z)\nabla_{\theta}\varphi(\pi^*(\theta,z) \mid z,\theta)\Big]$$
$$+ \nabla_{\theta\pi'}\varphi(\pi^*(\theta,z) \mid z,\theta)'\Bigg\{\mathbf{I}_{k+2} \otimes \bigg(J\big(\pi^*(\theta,z) \mid z,\theta\big)^{-1}\frac{1}{f_Z(z)}\big[\widehat{S}_N\big(\pi^*(\theta,z) \mid z,\theta\big) - \widehat{f}_{Z_N}(z)\pi^*(\theta,z)\big]\bigg)\Bigg\}$$
$$+ o_p(\mathrm{N}^{-1/2}) \qquad (\clubsuit\ 1)$$

for all $(\theta,z) \in \Theta \times \mathcal{Z}$.

Let $\underset{(4\times 2)}{\nabla_{\pi}\Big(J(\pi \mid z,\theta)^{-1}\Big)} = \nabla_{\pi}\mathrm{vec}\Big(J(\pi \mid z,\theta)^{-1}\Big)$ and $\underset{(4\times 2)}{\nabla_{\pi}\Big(\widehat{J}_N(\pi \mid z,\theta)^{-1}\Big)} = \nabla_{\pi}\mathrm{vec}\Big(\widehat{J}_N(\pi \mid z,\theta)^{-1}\Big)$ Using Lemma A.2 and the fact that $\pi^*(\theta,z) \in [0,1]^2$ (a compact set) for all $(\theta,z) \in \Theta \times \mathcal{Z}$ along with assumption (S3.2), we can use a mean-value approximation to show that

[29]As in all previous mean-value approximations, the fact that $\pi^*(\theta,z) \in [0,1]^2$ and $\widehat{\pi^*_N}(\theta,z) \in [0,1]^2$ for all for all $(\theta,z) \in \Theta \times \mathcal{Z}$ implies that all mean values $\widetilde{\pi^*_N}(\theta,z)$ are also in $[0,1]^2$ (a compact set) for all $(\theta,z) \in \Theta \times \mathcal{Z}$, which allows us to use Lemma A.2.

$\sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}} \left\|\nabla_{\pi}\left(\widehat{J}_N(\pi^*(\theta,z) \mid z,\theta)^{-1}\right) - \nabla_{\pi}\left(J(\pi^*(\theta,z) \mid z,\theta)^{-1}\right)\right\| = o_p(\mathrm{N}^{-1/4})$. By Theorem A.5 and Lemma A.7 we have: [30] $\sup_{\substack{z\in\mathcal{Z}\\ \theta\in\Theta}} \left\|\nabla_{\pi}\left(\widehat{J}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta)^{-1}\right) - \nabla_{\pi}\left(J(\pi^*(\theta,z) \mid z,\theta)^{-1}\right)\right\| = o_p(\mathrm{N}^{-1/4})$.
Using this result along with Lemmas A.6(B) and A.7 and doing a second-order approximation we get:

$$\widehat{J}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta)^{-1} - J(\pi^*(\theta,z) \mid z,\theta)^{-1} = \widehat{J}_N(\pi^*(\theta,z) \mid z,\theta)^{-1} - J(\pi^*(\theta,z) \mid z,\theta)^{-1}$$

$$+ \nabla_{\pi}\left(J(\pi^*(\theta,z) \mid z,\theta)^{-1}\right)'\left\{\mathbf{I}_2 \otimes \left(J\big(\pi^*(\theta,z) \mid z,\theta\big)^{-1}\frac{1}{f_{Z}(z)}\left[\widehat{S}_N\big(\pi^*(\theta,z) \mid z,\theta\big) - \widehat{f}_{Z_N}(z)\pi^*(\theta,z)\right]\right)\right\}$$

$$+ o_p(\mathrm{N}^{-1/2})$$

for all $(\theta,z) \in \Theta \times \mathcal{Z}$.

Now define:

$$\underset{(2\times 4)}{R(\pi \mid z,\theta)} = \frac{1}{d(\pi \mid z,\theta)^2}\begin{pmatrix} \alpha_1\alpha_2\delta_2(\pi_1 \mid z,\theta) & \alpha_1\alpha_2\delta_1(\pi_2 \mid z,\theta) & 1 & \alpha_1\alpha_2\delta_1^2(\pi_2 \mid z,\theta) \\ \alpha_1\alpha_2\delta_2^2(\pi_1 \mid z,\theta) & 1 & \alpha_1\alpha_2\delta_2(\pi_1 \mid z,\theta) & \alpha_1\alpha_2\delta_1(\pi_2 \mid z,\theta) \end{pmatrix}$$

Then, using (S3.2) and Lemma A.2 along with the fact that $\pi(\theta,z) \in [0,1]^2$, we have that a second order approximation for the term $\widehat{J}_N(\pi^*(\theta,z) \mid z,\theta)^{-1} - J(\pi^*(\theta,z) \mid z,\theta)^{-1}$ yields:

$$\widehat{J}_N(\pi^*(\theta,z) \mid z,\theta)^{-1} - J(\pi^*(\theta,z) \mid z,\theta)^{-1} =$$
$$R(\pi^*(\theta,z) \mid z,\theta)\left\{\mathbf{I}_2 \otimes \frac{1}{f_{Z}(z)}\left[\widehat{s}_N(\pi^*(\theta,z) \mid z,\theta) - \widehat{f}_{Z_N}(z)\delta(\pi^*(\theta,z) \mid z,\theta)\right]\right\} + o_p(\mathrm{N}^{-1/2})$$

for all $(\theta,z) \in \Theta \times \mathcal{Z}$.

Therefore, we finally get:

$$\widehat{J}_N(\widehat{\pi^*_N}(\theta,z) \mid z,\theta)^{-1} - J(\pi^*(\theta,z) \mid z,\theta)^{-1} =$$
$$R(\pi^*(\theta,z) \mid z,\theta)\left\{\mathbf{I}_2 \otimes \frac{1}{f_{Z}(z)}\left[\widehat{s}_N(\pi^*(\theta,z) \mid z,\theta) - \widehat{f}_{Z_N}(z)\delta(\pi^*(\theta,z) \mid z,\theta)\right]\right\}$$
$$+ \nabla_{\pi}\left(J(\pi^*(\theta,z) \mid z,\theta)^{-1}\right)'\left\{\mathbf{I}_2 \otimes \left(J\big(\pi^*(\theta,z) \mid z,\theta\big)^{-1}\frac{1}{f_{Z}(z)}\left[\widehat{S}_N\big(\pi^*(\theta,z) \mid z,\theta\big) - \widehat{f}_{Z_N}(z)\pi^*(\theta,z)\right]\right)\right\}$$
$$+ o_p(\mathrm{N}^{-1/2}) \qquad (\clubsuit\, 2)$$

[30]By the same argument as the one used in footnote 29, all the mean values satisfy: $\widetilde{\pi^*_N}(\theta,z) \in [0,1]^2$ for all $(\theta,z) \in \Theta \times \mathcal{Z}$ which not only allows us to apply Lemma A.2, but also assumption (S3.2) which is satisfied in $[0,1]^2$.

for all $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

We have

$$\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) = \widehat{J}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$$

$$\nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)$$

therefore

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) &= J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\Big[\nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\Big] \\
&+ \Big[\widehat{J}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\Big] \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \\
&+ \Big[\widehat{J}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\Big]\Big[\nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\Big]
\end{aligned}$$

Using (♣ $1-2$) along with Lemma A.2 and assumption (S3.2) we have:

$$\begin{aligned}
\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Bigg\{ \Bigg\| &\Big[\widehat{J}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\Big] \\
&\times \Big[\nabla_{\boldsymbol{\theta}} \widehat{\varphi}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)\Big] \Bigg\| \Bigg\} = o_p(\mathrm{N}^{-1/2})
\end{aligned}$$

therefore, using (♣ $1-2$) we get:

$$\begin{aligned}
&\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) = \\
&J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1}\Bigg\{ \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{A}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \nabla_{\boldsymbol{\theta}} \varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})\Big] \\
&+ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})' \Bigg[\mathbf{I}_{k+2} \otimes \Bigg(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big]\Bigg)\Bigg]\Bigg\} \\
&+ \Bigg\{ R(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}) \Bigg[\mathbf{I}_2 \otimes \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\big[\widehat{s}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \delta(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})\big]\Bigg] \\
&+ \nabla_{\boldsymbol{\pi}}\Big(J(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta})^{-1}\Big)' \Bigg[\mathbf{I}_2 \otimes \Bigg(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big]\Bigg)\Bigg]\Bigg\} \\
&\times \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) + o_p(\mathrm{N}^{-1/2}) \qquad (\clubsuit\ 3)
\end{aligned}$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Now let:

$$\begin{aligned}
\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) &= J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\widehat{A}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) \\
&+ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})'\Big[\mathbf{I}_{k+2} \otimes \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)\Big] \\
&+ \Big\{R(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})\Big[\mathbf{I}_2 \otimes \widehat{s}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})\Big] \qquad (\clubsuit\ 4) \\
&+ \nabla_{\boldsymbol{\pi}}\Big(J(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})^{-1}\Big)'\Big[\mathbf{I}_2 \otimes \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)\Big]\Big\}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)
\end{aligned}$$

and

$$\begin{aligned}
\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) &= J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}) \\
&+ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\varphi(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})'\Big[\mathbf{I}_{k+2} \otimes \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big)\Big] \\
&+ \Big\{R(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})\Big[\mathbf{I}_2 \otimes \delta\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big] \qquad (\clubsuit\ 5) \\
&+ \nabla_{\boldsymbol{\pi}}\Big(J(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta})^{-1}\Big)'\Big[\mathbf{I}_2 \otimes \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big)\Big]\Big\}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)
\end{aligned}$$

Then ($\clubsuit$ 3) becomes:

$$\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big] + o_p(\mathrm{N}^{-1/2})$$

for all $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$.

Lastly, note that by definition of these objects (see the proof of Lemma A.2), Lemma A.2 and Theorem A.5 and assumption (S3.2) we have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - f_{\boldsymbol{Z}}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\| = o_p(\mathrm{N}^{-1/4})$$

which completes the proof. □

We are now ready to prove Theorem 1.

Proof of Theorem 1

Recall throughout that $\boldsymbol{W} = (\boldsymbol{Y}', \boldsymbol{X}', \boldsymbol{Z}')'$. Now let us clarify the following notation:

$$\nabla_{\boldsymbol{\theta}} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}\big) = \text{Partial derivative of } \ell_{\mathcal{Z}} \text{ with respect to } \boldsymbol{\theta}\text{, with } \boldsymbol{\pi} \text{ constant.}$$

$$\nabla_{\boldsymbol{\pi}} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}\big) = \text{Partial derivative of } \ell_{\mathcal{Z}} \text{ with respect to } \boldsymbol{\pi}\text{, with } \boldsymbol{\theta} \text{ constant.}$$

$$\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}\big) = \text{Second partial derivative of } \ell_{\mathcal{Z}} \text{ with respect to } \boldsymbol{\theta}\text{, with } \boldsymbol{\pi} \text{ constant.}$$

$$\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}\big) = \text{Second partial derivative of } \ell_{\mathcal{Z}} \text{ with respect to } \boldsymbol{\pi}\text{, with } \boldsymbol{\theta} \text{ constant.}$$

$$\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}\big) = \text{Cross partial derivative of } \ell_{\mathcal{Z}} \text{ with respect to } \boldsymbol{\theta} \text{ (holding } \boldsymbol{\pi} \text{ constant) and } \boldsymbol{\pi} \text{ (holding } \boldsymbol{\theta} \text{ constant).}$$

From Lemma 4.1, we know that $\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})$ is an M times differentiable function of $\boldsymbol{\theta}$ and $\boldsymbol{Z}$ everywhere in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Let $\partial \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)/\partial \boldsymbol{\theta}$ and $\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'$ denote the total first and second partial derivatives of $\ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)$ with respect to $\boldsymbol{\theta}$. Note that $\partial \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)/\partial \boldsymbol{\theta}$ is the score and $\Im_{\mathcal{Z}} = -E\big[\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)/\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'\big]$ is the information matrix of the trimmed log-likelihood $\ell_{\mathcal{Z}}$. We have:

$$\underset{(k+2)\times 1}{\frac{\partial \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)}{\partial \boldsymbol{\theta}}} = \nabla_{\boldsymbol{\theta}} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big) + \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})' \nabla_{\boldsymbol{\pi}} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)$$

$$\underset{(k+2)\times(k+2)}{\frac{\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}} = \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big) + \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big) \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})$$

$$+ \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})' \big[\nabla_{\boldsymbol{\pi}} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big) \otimes \boldsymbol{I}_{(k+2)}\big] + \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})' \Big\{\nabla_{\boldsymbol{\pi}\boldsymbol{\theta}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big)$$

$$+ \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\big) \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z})\Big\}$$

where $\boldsymbol{I}_{(k+2)}$ is a $(k+2) \times (k+2)$ identity matrix.

It is easy to see that $E\big[\nabla_{\boldsymbol{\pi}} \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) \mid \boldsymbol{X}, \boldsymbol{Z}\big] = \boldsymbol{0}$ and therefore the trimmed information matrix $\Im_{\mathcal{Z}}$ is given by:

$$\Im_{\mathcal{Z}} = -E\left[\frac{\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right] =$$

$$- E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})$$

$$+ \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})' \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)' + \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})' \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\Big]$$

It is also easy to show that:

$$E\big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) \mid \boldsymbol{X},\boldsymbol{Z}\big] = -E\big[\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)' \mid \boldsymbol{X},\boldsymbol{Z}\big]$$

$$E\big[\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) \mid \boldsymbol{X},\boldsymbol{Z}\big] = -E\big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)' \mid \boldsymbol{X},\boldsymbol{Z}\big]$$

$$E\big[\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) \mid \boldsymbol{X},\boldsymbol{Z}\big] = -E\big[\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)' \mid \boldsymbol{X},\boldsymbol{Z}\big]$$

therefore the expression given above for $\Im_{\boldsymbol{\mathcal{Z}}}$ can be simplified to:

$$\begin{aligned}\Im_{\boldsymbol{\mathcal{Z}}} &= -E\Big[\big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\big\} \\ &\qquad \times \big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\big\}'\Big] \\ &= -E\left[\frac{\partial\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}} \times \frac{\partial\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}'\right]\end{aligned}$$

which implies that the trimmed log-likelihood $\ell_{\boldsymbol{\mathcal{Z}}}$ satisfies an information identity result. Now let $\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big)/\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'$ denote the partial derivative of the score vector with respect to $\boldsymbol{\pi}$. Then, using iterated expectations once again it is easy to show that:

$$\begin{aligned}\overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z}) &\equiv E\left[\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'}\Big|\ \boldsymbol{Z}\right] \\ &= E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\ \Big|\boldsymbol{Z}\Big]\end{aligned}$$

We are now ready to show consistency of $\widehat{\boldsymbol{\theta}}$:

Proof of Theorem 1(A):

From Lemma 4.1 and assumption (S3.2), $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})$ is continuous in $\boldsymbol{\Theta}\times\boldsymbol{\mathcal{Z}}$. Combining this with the continuity of the linear function $\boldsymbol{X}'\boldsymbol{\beta}+\alpha\pi$ and assumption (S1.3), then $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ is continuous in $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$. By assumptions (S2.3) and (S3), the set $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$ is compact and therefore the continuity of $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ is uniform in $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$. In addition, from Lemma 4.1 we know that there exists $b\in(0,1)$ such that $\inf\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big)>b$ and $1-\sup\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\Big)>b$.Now, take any $\boldsymbol{w}\in\big\{0,1\big\}\times\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}$ and any $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ with the corresponding $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})$. Then, by uniform continuity we have that for all $M>0$ there exists $\delta>0$ such that $\boldsymbol{\pi}\in[0,1]^2$ and $\Big\|\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\pi}\Big\|<\delta$ imply $\Big\|\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big)-\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi}\big)\Big\|<M$.

Now let $\overline{\delta} = \min\left\{\delta\ ,\ b\right\}$. Then we have $\overline{\delta} > 0$ and using Lemma 4.2(A) we have that for all $\varepsilon > 0$, there exists $N_{\overline{\delta}}$ such that $N > N_{\overline{\delta}}$ implies:

$$\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\right\| > \widetilde{\delta}\right\} < \varepsilon$$

Therefore, $N > N_{\overline{\delta}}$ implies

$$\Pr\left\{\sup_{\substack{\boldsymbol{w}\in\{0,1\}\times\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\big) - \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})\big)\right\| \geq M\right\} < \varepsilon$$

and consequently:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big) - \frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}_n)\big)\right| \xrightarrow{p} 0$$

From assumption (S4.3), the sample is iid. As we mentioned above, Lemma 4.1 and the continuity of the linear function $\boldsymbol{\beta}'\boldsymbol{X} + \alpha\pi$, imply that $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ is a continuous function at each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ with probability one. By (S3.1), $\boldsymbol{\Theta}$ is compact. We also know that $\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z}) \in [0,1]^2$ (a compact set) for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and all $\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}$. Compactness of $\{0,1\}\times\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times[0,1]$ implies that there exists $\overline{\ell}$ such that $\left|\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\right| < \overline{\ell}$ with probability one. These properties are sufficient to satisfy the assumptions of Lemma 2.4 in Newey and McFadden (1994) (dominated uniform convergence theorem) and imply that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}_n)\big) - E\big[\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]\right| = o_p(1)$$

These results together imply that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big) - E\big[\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]\right| = o_p(1)$$

From Lemma 4.4 we know that $E\big[\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]$ is uniquely maximized at $\boldsymbol{\theta}_0$. By Lemma 4.1, we know that $E\big[\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\big]$ is continuous. The result immediately above showed that $\frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)$ converges in probability to $E\big[\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]$ uniformly in $\boldsymbol{\Theta}$. Since $\widehat{\boldsymbol{\theta}}$ maximizes $\frac{1}{N}\sum_{n=1}^{N}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widehat{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)$ in $\boldsymbol{\Theta}$, all the conditions of Theorem 2.1 in Newey and McFadden are met and therefore $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$. We next prove part (B).

Proof of Theorem 1(B):

Before proceeding, note that the trimming index $\mathbb{1}\{\boldsymbol{Z} \in \boldsymbol{\mathcal{Z}}\}$ does not depend on $\boldsymbol{\theta}$. Then, using assumption (S1.3), compactness of $\boldsymbol{\Theta}$ and Lemma 4.1, we have that $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ is an M times differentiable function of $\boldsymbol{\theta}$ with bounded M derivatives. We also argued previously -see the discussion following assumption (S3)- that $\text{boundary}\big(\boldsymbol{\mathcal{Z}}\big) = \boldsymbol{\mathcal{Z}} \cap \text{ cl}\big(\boldsymbol{\mathcal{Z}}^c\big)$ has Lebesgue measure zero in $\mathbb{R}^L$. Since $\boldsymbol{Z}$ is continuously distributed ($\boldsymbol{Z}$ is absolutely continuous with respect to Lebesgue measure), we have $\Pr\{\boldsymbol{Z} \in \text{boundary}\big(\boldsymbol{\mathcal{Z}}\big)\} = 0$. Therefore, using once again assumption (S1.3), compactness of $\boldsymbol{\Theta}$ and Lemma 4.1, we have that with probability one, $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ is also an M times differentiable function of $\boldsymbol{Z}$ with bounded M derivatives. We now proceed to the proof: As we did in the proof of Lemma A.5, take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and define the indicator variable:

$$\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) = \mathbb{1}\Big\{\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \in (0,1)^2 \quad \text{and} \quad \widehat{d}_N\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) \neq 0\Big\}$$

we showed previously that $\Pr\left\{\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \big|\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) - 1\big| = 1\right\} \to 0$. As we outlined above, $\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) = 1$ implies that $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z})$ is an M times differentiable function of $\boldsymbol{\theta}$ and $\boldsymbol{z}$. Now, note that for each $\boldsymbol{z}_n \in \{\boldsymbol{z}_n\}_{n=1}^N$: $\mathbb{1}\big\{\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}\big\}\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big|\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}_n) - 1\big| = 1$ only if $\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \big|\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}) - 1\big| = 1$. Consequently, we have: $\Pr\Big\{\mathbb{1}\{\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}\}\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \big|\overline{1}_N(\boldsymbol{\theta}, \boldsymbol{z}_n) - 1\big| = 1 \text{ for at least some } \boldsymbol{z}_n \in \{\boldsymbol{z}_n\}_{n=1}^N\Big\} \to 0$. Therefore, with probability approaching one the estimator $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ satisfies the first order conditions:

$$\frac{1}{N}\sum_{n=1}^{N}\Big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}_N^*}(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}_N^*}(\widehat{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)\Big\} = \mathbf{0}$$

and $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ is an M times differentiable function of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and for all $\boldsymbol{z}_n$ (since $\boldsymbol{z}_n\in\boldsymbol{\mathcal{Z}}$ for all $\boldsymbol{z}_n$). A first order Taylor series approximation for $\widehat{\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_0$ yields:

$$-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)= \\ \frac{1}{N}\sum_{n=1}^{N}\Big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big\} \tag{5}$$

with $\widetilde{\boldsymbol{\theta}}$ between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ and:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}= \\ \frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n) \\ +\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}_n)'\big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\otimes\boldsymbol{I}_{(k+2)}\big]+\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)'\Big\{\nabla_{\boldsymbol{\pi}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big) \\ +\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\Big\}\Big]$$

where $\boldsymbol{I}_{(k+2)}$ is a $(k+2)\times(k+2)$ identity matrix.

We have:

$$\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-E\Big[\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]\right\| \\ \leq\sup_n\left\|\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| \\ +\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-E\Big[\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]\right\|$$

Lemma 4.1(A), assumption (S1.3) and the compactness of $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$ imply that the functions $\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$, $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$, $\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ and $\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\big)$ are all uniformly continuous in $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$. Since $\widetilde{\boldsymbol{\theta}}\in\boldsymbol{\Theta}$ then using Lemma 4.2(A) and taking the same steps as above we get: $\sup_n\Big\|\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)-\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\Big\|=o_p(1)$, $\sup_n\Big\|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)-\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\Big\|=o_p(1)$, $\sup_n\Big\|\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)-\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\Big\|=o_p(1)$, and $\sup_n\Big\|\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)-\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)\Big\|=o_p(1)$. The results in Lemma 4.2(B) and the trimming index $\mathbb{1}\{\boldsymbol{z}_n\in\boldsymbol{\mathcal{Z}}\}$ imply that

$\sup_n \left\| \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n) \right\| = o_p(1)$ and $\sup_n \left\| \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n) \right\| = o_p(1)$. These results together imply:

$$\sup_n \left\| \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| = o_p(1)$$

$\widetilde{\boldsymbol{\theta}}$ is intermediate between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Therefore $\widetilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$. By the same argument used above and the fact that $\boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n) \in [0,1]^2$ for all $\boldsymbol{z}_n$, we get that $\left\| \partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}, \boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}_n)\big) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \right\|$ is bounded with probability one for all $\boldsymbol{w}_n$, all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By Lemma 4.1, it is also a continuous function everywhere in $\boldsymbol{\Theta}$. Consequently, $E\left[\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \right]$ is continuous and bounded. Once again using Lemma 2.4 in Newey and McFadden, we get:

$$\left\| \frac{1}{N} \sum_{n=1}^{N} \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \boldsymbol{\pi}^*(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - E\left[\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\| \xrightarrow{p} 0$$

and consequently:

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xrightarrow{p} E\left[\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = -\Im_{\boldsymbol{\mathcal{Z}}} \tag{6}$$

Next we examine the terms in the right hand side of (5). A second order Taylor approximation for the first term yields:

$$\begin{aligned}
&\frac{1}{N} \sum_{n=1}^{N} \nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) = \frac{1}{N} \sum_{n=1}^{N} \nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \\
&+ \frac{1}{N} \sum_{n=1}^{N} \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \\
&+ \frac{1}{2N} \sum_{n=1}^{N} \Bigg[\left\{ \boldsymbol{I}_{(k+2)} \otimes \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \right) \right\}' \nabla_{\boldsymbol{\pi}} \text{vec} \left\{ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \right\} \\
&\qquad \times \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \right) \Bigg]
\end{aligned}$$

with each $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)$ between $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)$. We have $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \in [0,1]^2$ (a compact set) for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. By assumptions (S2.3) and (S3), $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$ is a compact set. By assumption (S1.3), $G_1(\cdot)$ and $G_2(\cdot)$ are $\mathcal{C}^{M+2}$ functions, with bounded $M+2$ derivatives. These facts imply that $\sup_n \left\| \nabla_{\boldsymbol{\pi}} \text{vec} \left\{ \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \right\} \right\|$ is bounded with probability one. Combining this with Lemma 4.2(A) and the fact that $\boldsymbol{z}_n \in \mathcal{Z}$ for all

n, we get:

$$\sup_n \left\| \frac{1}{N}\sum_{n=1}^{N}\left[\left\{\boldsymbol{I}_{(k+2)} \otimes \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right\}' \nabla_{\boldsymbol{\pi}} \text{vec}\left\{\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\right\} \times \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right]\right\| = o_p(\mathrm{N}^{-1/2})$$

and consequently:

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) &= \frac{1}{N}\sum_{n=1}^{N}\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) \\ &\quad \frac{1}{N}\sum_{n=1}^{N}\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) + o_p(\mathrm{N}^{-1/2})\end{aligned} \tag{7}$$

We now turn to the second term on the right hand side of (5). First, note that a second-order Taylor approximation yields:

$$\begin{aligned}&\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) = \\ &\quad \nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) \\ &\quad + \frac{1}{2}\left[\left\{\boldsymbol{I}_{(k+2)} \otimes \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right\}' \nabla_{\boldsymbol{\pi}}\text{vec}\left\{\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\right\} \right. \\ &\quad \left. \times \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right]\end{aligned}$$

where $\boldsymbol{I}_2$ is a 2×2 identity matrix and each $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)$ between $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)$. We have $\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) \in [0,1]^2$ (a compact set) for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. By assumptions (S2.3) and (S3), $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$ is a compact set. By assumption (S1.3), $G_1(\cdot)$ and $G_2(\cdot)$ are $\mathcal{C}^{M+2}$ functions, with bounded $M+2$ derivatives. These facts imply that $\sup_n \left\|\nabla_{\boldsymbol{\pi}}\text{vec}\left\{\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\right\}\right\|$ is bounded with probability one. Combining this with Lemma 4.2(A) and the fact that $\boldsymbol{z}_n \in \mathcal{Z}$ for all n, we get:

$$\sup_n \left\|\left\{\boldsymbol{I}_{(k+2)} \otimes \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right\}' \nabla_{\boldsymbol{\pi}}\text{vec}\left\{\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\right\} \times \left(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\right)\right\| = o_p(\mathrm{N}^{-1/2})$$

and therefore:

$$\begin{aligned}\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) &= \nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) \\ &\quad + \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) + o_p(\mathrm{N}^{-1/2}) \quad \text{for all } n\end{aligned}$$

Using this result, adding and subtracting $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)$ we can express the second term in the right hand side of (5) as:

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) = \frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$
$$+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)$$
$$+\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$
$$+\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)\Big]+o_p(\mathrm{N}^{-1/2})$$

We have $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \in [0,1]^2$ (a compact set) for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. By assumptions (S2.3) and (S3), $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$ is a compact set. By assumption (S1.3), $G_1(\cdot)$ and $G_2(\cdot)$ are $\mathcal{C}^{M+2}$ functions, with bounded $M+2$ derivatives. These facts imply that $\sup_n \left\|\nabla_{\boldsymbol{\pi}}\mathrm{vec}\left\{\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\right\}\right\|$ is bounded with probability one. Combining this with Lemma 4.2(A)-(B) and the fact that $\boldsymbol{z}_n \in \mathcal{Z}$ for all n, we get:

$$\sup_n\left\|\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)\right\|$$
$$=o_p(\mathrm{N}^{-1/2})$$

and therefore:

$$\frac{1}{N}\sum_{n=1}^{N}\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) = \frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$
$$+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)$$
$$+\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big]+o_p(\mathrm{N}^{-1/2}) \tag{8}$$

Define

$$\underset{(k+2)\times 2}{D_{\boldsymbol{\mathcal{Z}}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) \equiv \Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big]$$

Then, using Equations (7) and (8) we get:

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big]= \\
&\qquad\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big] \\
&\qquad+\frac{1}{N}\sum_{n=1}^{N}\Big[D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big) \\
&\qquad+\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big]+o_p(\mathrm{N}^{-1/2})
\end{aligned} \tag{9}$$

Using Lemma A.6 and the fact that $\boldsymbol{z}_n\in\boldsymbol{\mathcal{Z}}$ for all n, we have:

$$\begin{aligned}
&\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)= \\
&J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)^{-1}\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)-\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n)\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big]+o_p(\mathrm{N}^{-1/2})\quad\text{for all } n
\end{aligned}$$

We have: $E[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}]=\Big(G_1(\boldsymbol{X}_1'\boldsymbol{\beta}_{1_0}+\alpha_{1_0}\pi_2^*(\boldsymbol{\theta}_0,\boldsymbol{Z}))\ ,\ G_2(\boldsymbol{X}_2'\boldsymbol{\beta}_{2_0}+\alpha_{2_0}\pi_1^*(\boldsymbol{\theta}_0,\boldsymbol{Z}))\Big)'$. By definition, we also have $E[\boldsymbol{Y}\mid\boldsymbol{Z}]=\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})$. Therefore, by definition (see Lemma A.6) we can express:

$$\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)=\frac{1}{Nh_N^L}\sum_{m=1}^{N}E[\boldsymbol{Y}\mid\boldsymbol{x}_m,\boldsymbol{z}_n]K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)$$

and using the result of Lemma A.6 cited above, we get:

$$\begin{aligned}
&\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)-\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n)\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big]= \\
&\qquad\qquad\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\frac{1}{Nh_N^L}\sum_{m=1}^{N}\Big(E[\boldsymbol{Y}\mid\boldsymbol{x}_m,\boldsymbol{z}_n]-E[\boldsymbol{Y}\mid\boldsymbol{z}_n]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)
\end{aligned}$$

Define $B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\equiv D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)^{-1}$. Then using the result immediately above, we get:

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)= \\
&\frac{1}{N^2h_N^L}\sum_{n=1}^{N}B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\sum_{m=1}^{N}\Big(E[\boldsymbol{Y}\mid\boldsymbol{x}_m,\boldsymbol{z}_n]-E[\boldsymbol{Y}\mid\boldsymbol{z}_n]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n) \\
&=\frac{1}{N^2h_N^L}\sum_{n=1}^{N}\sum_{m\neq n}B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(E[\boldsymbol{Y}\mid\boldsymbol{x}_m,\boldsymbol{z}_n]-E[\boldsymbol{Y}\mid\boldsymbol{z}_n]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n) \\
&\quad+\frac{K(0)}{Nh_N^L}\frac{1}{N}\sum_{n=1}^{N}B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(E[\boldsymbol{Y}\mid\boldsymbol{x}_n,\boldsymbol{z}_n]-E[\boldsymbol{Y}\mid\boldsymbol{z}_n]\Big)
\end{aligned}$$

We have $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \in [0,1]^2$ for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. Consequently, using assumption (S3.2), $\left\| J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \mid \boldsymbol{z}_n, \boldsymbol{\theta}_0\big)^{-1} \right\|$ is uniformly bounded for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. By (S3.2), there exists a $b > 0$ such that $\left| f_{\boldsymbol{Z}}(\boldsymbol{z}_n)^{-1} \right| < b$ for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$. By an argument parallel to the one used in the paragraph previous to Equation (8), assumption (S1.3) implies that $G_1(\cdot)$ and $G_2(\cdot)$ are $\mathcal{C}^{M+2}$ functions, with bounded $M+2$ derivatives. This, along with the fact that $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \in [0,1]^2$ (a compact set) for all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$ implies that $\sup_n \left\| D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \right\|$ is bounded with probability one. By definition of our trimmed log-likelihood function, we have: $D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \mid \boldsymbol{z}_n, \boldsymbol{\theta}_0\big)^{-1} \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} = \mathbf{0}$ whenever $\boldsymbol{z}_n \notin \boldsymbol{\mathcal{Z}}$ (since $\boldsymbol{z}_n \notin \boldsymbol{\mathcal{Z}}$ implies $D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) = \mathbf{0}$). Therefore, there exists $\boldsymbol{C} > 0$ such that:

$$\sup_n \left\| B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \right\| \leq \boldsymbol{C} \quad \text{w.p.1} \tag{10}$$

From now on, to simplify the notation we will denote $B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \equiv B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n\big)$, $D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) \equiv D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n\big)$ -recall that $\boldsymbol{w}_n = (\boldsymbol{y}_n', \boldsymbol{x}_n', \boldsymbol{z}_n')'$ - and $J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \mid \boldsymbol{z}_n, \boldsymbol{\theta}_0\big) \equiv J_0(\boldsymbol{z}_n)$. Then:

$$\frac{1}{N} \sum_{n=1}^{N} B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n\big) \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big] \Big)$$
$$\xrightarrow{p} E\left[B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{W}) \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{Z})} \Big(E\big[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{Z}\big] \Big) \right]$$

which exists and is finite by (10). Now, by assumption (S4.2.i) we have $N^{1/2}/(Nh_N^L) \to 0$. By (S4.1), we know that $K(0)$ is finite and therefore:

$$\frac{K(0)}{Nh_N^L} \frac{1}{N} \sum_{n=1}^{N} B_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n\big) \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big] \Big) = o_p(\mathrm{N}^{-1/2})$$

Consequently:

$$\frac{1}{N} \sum_{n=1}^{N} D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n\big) \Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n) \Big) =$$
$$\frac{1}{N^2 h_N^L} \sum_{n=1}^{N} \sum_{m \neq n} B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{w}_n) \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big] \Big) K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) + o_p(\mathrm{N}^{-1/2})$$

Before proceeding, let us write the basic Central Limit Theorem of the so-called U-statistics. This is a well-known result widely used in semiparametric and nonparametric estimation problems. We follow the results of Powell, Stock and Stoker (1989), see also Appendix A.2 in Pagan and Ullah (1999).

Central Limit Theorem for U-statistics: Consider a general second-order U-statistic of the form

$$U_N = \binom{N}{2}^{-1} \sum_{1\leq n<m\leq N} T_N(\boldsymbol{w}_n, \boldsymbol{w}_m)$$

where $\boldsymbol{w}_n$, n=1,...,N is an i.i.d random vector and T_n satisfies $T_N(\boldsymbol{w}_n, \boldsymbol{w}_m) = T_N(\boldsymbol{w}_m, \boldsymbol{w}_n)$. Now define $t_N(\boldsymbol{w}_n) = E[T_N(\boldsymbol{w}_n, \boldsymbol{w}_m) \mid \boldsymbol{w}_n]$ and

$$\widehat{U}_N = E\big[t_N(\boldsymbol{w}_n)\big] + \frac{2}{N}\sum_{n=1}^{N}\Big\{t_N(\boldsymbol{w}_n) - E\big[t_N(\boldsymbol{w}_n)\big]\Big\}$$

$\widehat{U}_N$ is called the "projection" of the statistic U_n.
Then $\sqrt{N}(U_N - \widehat{U}_N) = o_p(1)$ if $E\big[\big\|T_N(\boldsymbol{w}_m, \boldsymbol{w}_n)\big\|^2\big] = o(N)$ □

Let

$$\begin{aligned} T_N(\boldsymbol{w}_n, \boldsymbol{w}_m) = &\frac{B_{\boldsymbol{Z}}(\boldsymbol{w}_n)}{h_N^L f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) \\ &+ \frac{B_{\boldsymbol{Z}}(\boldsymbol{w}_m)}{h_N^L f_{\boldsymbol{Z}}(\boldsymbol{z}_m)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_m] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_m\big]\Big)K_h(\boldsymbol{z}_n - \boldsymbol{z}_m) \end{aligned}$$

and let $U_N = \binom{N}{2}^{-1}\sum_{1\leq n<m\leq N} T_N(\boldsymbol{w}_n, \boldsymbol{w}_m)$. By symmetry of $K(\cdot)$ (assumption (S4.1)) we have $K_h(\boldsymbol{z}_n - \boldsymbol{z}_m) = K_h(\boldsymbol{z}_m - \boldsymbol{z}_n)$ for all $\boldsymbol{z}_n, \boldsymbol{z}_m$. Then, we can re-express:

$$\frac{1}{N}\sum_{n=1}^{N} D_{\boldsymbol{Z}}(\boldsymbol{w}_n)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\Big) = \left(\frac{N-1}{2N}\right)U_N + o_p(\mathrm{N}^{-1/2})$$

we will now determine $t_N(\boldsymbol{w}_n) = E\big[T_N(\boldsymbol{w}_n, \boldsymbol{w}_m) \mid \boldsymbol{w}_n\big]$. We begin with the first term on the right hand side of $T_N(\boldsymbol{w}_n, \boldsymbol{w}_m)$. By the iid nature of the sample (assumption (S4.3)), we have:

$$\begin{aligned} &E\left[\frac{B_{\boldsymbol{Z}}(\boldsymbol{w}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) \,\Big|\boldsymbol{w}_n\right] \\ &\quad= \frac{B_{\boldsymbol{Z}}(\boldsymbol{w}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\int\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{u}, \boldsymbol{z}_n] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)\frac{1}{h_N^L}\int K\left(\frac{\boldsymbol{v} - \boldsymbol{z}_n}{h_N}\right)f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{v})d\boldsymbol{v}d\boldsymbol{u} \\ &\quad= \frac{B_{\boldsymbol{Z}}(\boldsymbol{w}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\int\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{u}, \boldsymbol{z}_n] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)\int K(\boldsymbol{\Psi})f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n + h_N\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u} \end{aligned}$$

with $\boldsymbol{\Psi} \equiv (\boldsymbol{v} - \boldsymbol{z}_n)/h_N$. Define:

$$Q_i \equiv \{(q_1, \ldots, q_L) \in \mathbb{N}^L : q_1 + \cdots + q_L = i\} \quad \text{and} \quad \Gamma_i(\boldsymbol{u}, \boldsymbol{z}) = \sum_{Q_i} \frac{\partial^i f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z})}{\partial z_1^{q_1} \cdots \partial z_L^{q_L}}$$

then by (S2.2) the following Taylor series approximation is valid:

$$\begin{aligned}
&\int K(\boldsymbol{\Psi}) f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n + h_N \boldsymbol{\Psi}) d\boldsymbol{\Psi} = \\
&\qquad f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n) \int K(\boldsymbol{\Psi}) d\boldsymbol{\Psi} + \sum_{i=1}^{M-1} (-1)^i \frac{h_N^i}{i!} \Gamma_i(\boldsymbol{u}, \boldsymbol{z}_n) \sum_{Q_i} \int \Psi_1^{q_1} \cdots \Psi_L^{q_L} K(\boldsymbol{\Psi}) d\boldsymbol{\Psi} \\
&\qquad + (-1)^M \frac{h_N^M}{M!} \int \sum_{Q_M} (\Psi_1^{q_1} \cdots \Psi_L^{q_L}) \Gamma_M(\boldsymbol{u}, \boldsymbol{z}_n + h_N^* \boldsymbol{\Psi}) K(\boldsymbol{\Psi}) d\boldsymbol{\Psi}
\end{aligned}$$

where h_N^* is between h_N and zero. By (S2.2) there exists a $\overline{C}_1 < \infty$ such that $\|\Gamma_i(\boldsymbol{u}, \boldsymbol{z})\| < \overline{C}_1$ for all $(\boldsymbol{u}, \boldsymbol{z}) \in \mathbb{R}^{k+L}$ and all $i \in \{1, \ldots, M\}$. This, along with (S4.1) implies that the first term on the right hand side is $f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n)$, the second term is zero and the third term is bounded for all $\boldsymbol{z}_n$. By (S4.2.ii) we have $h_N^M = o(\mathrm{N}^{-1/2})$ and therefore $\int K(\boldsymbol{\Psi}) f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n + h_N \boldsymbol{\Psi}) d\boldsymbol{\Psi} = f_{\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n) + o_p(\mathrm{N}^{-1/2})$ for each $\boldsymbol{z}_n$, which yields:

$$\begin{aligned}
&E\left[\frac{B_{\mathcal{Z}}(\boldsymbol{w}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \Big(E[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n] - E[\boldsymbol{Y} \mid \boldsymbol{z}_n]\Big) K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) \,\Big|\boldsymbol{w}_n\right] \\
&\qquad = \frac{B_{\mathcal{Z}}(\boldsymbol{w}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \int \Big(E[\boldsymbol{Y} \mid \boldsymbol{u}, \boldsymbol{z}_n] - E[\boldsymbol{Y} \mid \boldsymbol{z}_n]\Big) f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n) d\boldsymbol{u} + o_p(\mathrm{N}^{-1/2}) \\
&\qquad = o_p(\mathrm{N}^{-1/2})
\end{aligned} \tag{11}$$

where the last equality follows from the fact that

$$\int E[\boldsymbol{Y} \mid \boldsymbol{u}, \boldsymbol{z}_n] \frac{f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} d\boldsymbol{u} = E\Big[E[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{z}_n] \mid \boldsymbol{z}_n\Big] = E[\boldsymbol{Y} \mid \boldsymbol{z}_n] \text{ and } \int f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n) d\boldsymbol{u} = f_{\boldsymbol{Z}}(\boldsymbol{z}_n)$$

which together imply that $\displaystyle \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \int \Big(E[\boldsymbol{Y} \mid \boldsymbol{u}, \boldsymbol{z}_n] - E[\boldsymbol{Y} \mid \boldsymbol{z}_n]\Big) f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u}, \boldsymbol{z}_n) d\boldsymbol{u} = 0$

Before turning to the second term on the right hand side of $T_N(\boldsymbol{w}_n, \boldsymbol{w}_m)$, let us define $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z}) = E\Big[D_{\mathcal{Z}}(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})) \mid \boldsymbol{Z}\Big]$. Then, by definition (see the line immediately after Equation (8)) we have:

$$\overline{D}_{\mathcal{Z}}(\boldsymbol{Z}) = E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})) \big| \boldsymbol{Z}\Big] + \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})' E\Big[\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'} \ell_{\mathcal{Z}}(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})) \big| \boldsymbol{Z}\Big]$$

then, using assumptions (S1.3), (S2), (S3.2) and Lemma 4.1, we have that w.p.1, $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})$ is an M times differentiable function of $\boldsymbol{Z}$ with bounded M derivatives. Finally, define

$\overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z}) = \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})J_0(\boldsymbol{Z})^{-1}$. Using the result immediately cited above, along with Lemma 4.1 and (S3.2), we get that $\overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})$ is also an M times differentiable function of $\boldsymbol{Z}$ with bounded M derivatives. Using iterated expectations and the iid nature of the data, we have:

$$
\begin{aligned}
E\Big[\frac{B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{w}_m)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_m\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_m\big]\Big)K_h(\boldsymbol{z}_m - \boldsymbol{z}_n)\ \Big| \boldsymbol{w}_n\Big]& \\
&= \int \frac{\overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{v})}{h_N^L f_{\boldsymbol{Z}}(\boldsymbol{v})}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{v}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{v}\big]\Big)K\Big(\frac{\boldsymbol{v} - \boldsymbol{z}_n}{h_N}\Big)f_{\boldsymbol{Z}}(\boldsymbol{v})d\boldsymbol{v} \\
&= \int \frac{\overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{v})}{h_N^L}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{v}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{v}\big]\Big)K\Big(\frac{\boldsymbol{v} - \boldsymbol{z}_n}{h_N}\Big)d\boldsymbol{v} \\
&= \int \overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n + h_N\boldsymbol{\Psi})\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n + h_N\boldsymbol{\Psi}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n + h_N\boldsymbol{\Psi}\big]\Big)K\big(\boldsymbol{\Psi}\big)d\boldsymbol{\Psi}
\end{aligned}
$$

Now denote $A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{X}, \boldsymbol{Z}) = \overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{Z}\big]\Big)$. As we did above, let:

$$
Q_i \equiv \{(q_1, \ldots, q_L) \in \mathbb{N}^L : q_1 + \cdots + q_L = i\} \quad \text{and now define } \Delta_i(\boldsymbol{x}, \boldsymbol{z}) = \sum_{Q_i} \frac{\partial^i A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}, \boldsymbol{z})}{\partial z_1^{q_1} \cdots \partial z_L^{q_L}}
$$

Once again, using Lemma 4.1 and assumption (S1.3), along with the result concerning $\overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{Z})$ mentioned immediately above, we have that $A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{X}, \boldsymbol{Z})$ is also an M times differentiable function of $\boldsymbol{Z}$ with bounded M derivatives w.p.1. Using this result along with assumption (S2.2), the following Taylor series approximation is valid:

$$
\begin{aligned}
\int A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n + h_N\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi} =& \\
& A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n)\int K(\boldsymbol{\Psi})d\boldsymbol{\Psi} + \sum_{i=1}^{M-1}(-1)^i\frac{h_N^i}{i!}\Delta_i(\boldsymbol{x}_n, \boldsymbol{z}_n)\sum_{Q_i}\int \Psi_1^{q_1}\cdots\Psi_L^{q_L}K(\boldsymbol{\Psi})d\boldsymbol{\Psi} \\
& + (-1)^M\frac{h_N^M}{M!}\int\sum_{Q_M}(\Psi_1^{q_1}\cdots\Psi_L^{q_L})\Delta_M(\boldsymbol{x}_n, \boldsymbol{z}_n + h_N^*\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi}
\end{aligned}
$$

where h_N^* is between h_N and zero. Now, because $A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{X}, \boldsymbol{Z})$ is an M times differentiable function of $\boldsymbol{Z}$ with bounded M derivatives w.p.1, there exists a $\overline{C}_2 < \infty$ such that $\|\Delta_i(\boldsymbol{x}, \boldsymbol{z})\| < \overline{C}_2$ w.p.1 for all $(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{k+L}$ and all $i \in \{1, \ldots, M\}$. This, along with (S4.1) implies that the first term on the right hand side is $A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n)$, the second term is zero and the third term is bounded for all n. By (S4.2.ii) we have $h_N^M = o(\mathrm{N}^{-1/2})$ and therefore $\int A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n +$

$h_N\boldsymbol{\Psi})K(\boldsymbol{\Psi})d\boldsymbol{\Psi} = A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n) + o_p(\mathrm{N}^{-1/2})$ for each n, which yields:

$$\begin{aligned} E\Big[\frac{B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{w}_m)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_m\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_m\big]\Big)K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) \,\Big|\boldsymbol{w}_n\Big] \\ = A_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{x}_n, \boldsymbol{z}_n) + o_p(\mathrm{N}^{-1/2}) \\ = \overline{B}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2}) \\ = \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2}) \end{aligned} \tag{12}$$

combining Equations (11-12) we get:

$$\begin{aligned} t_N(\boldsymbol{w}_n) \equiv E\big[T_N(\boldsymbol{w}_n, \boldsymbol{w}_m) \mid \boldsymbol{w}_n\big] \\ = \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

Note that $E\big[t_N(\boldsymbol{w}_n) \mid \boldsymbol{z}_n\big] = 0$, which implies (by iterated expectations) that $E\big[t_N(\boldsymbol{w}_n)\big] = 0$. Using Equation (10) along with Lemma 4.1 and assumption (S4.1) -boundedness of $K(\cdot)$- we can show that:

$$\begin{aligned} E\Big\|\frac{B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{w}_n)}{h_N^L f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_m, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m - \boldsymbol{z}_n) \\ + \frac{B_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{w}_m)}{h_N^L f_{\boldsymbol{Z}}(\boldsymbol{z}_m)}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_m\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_m\big]\Big)K_h(\boldsymbol{z}_n - \boldsymbol{z}_m)\Big\|^2 = \frac{1}{h_N^L}O(1) \end{aligned}$$

or equivalently:

$$E\Big\|T_N(\boldsymbol{w}_n, \boldsymbol{w}_m)\Big\|^2 = \frac{1}{h_N^L}O(1) = o(N)$$

where the last equality follows from assumption (S4.2.i), which implies $Nh_N^L \to \infty$. Therefore, the condition of the CLT for U-statistics is satisfied and combining Equations (11-12) we have:

$$U_N = \frac{2}{N}\sum_{n=1}^{N}\overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2})$$

therefore:

$$\begin{aligned} \frac{1}{N}\sum_{n=1}^{N}D_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)\Big(\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\Big) \\ = \Big(\frac{N-1}{2N}\Big)U_N + o_p(\mathrm{N}^{-1/2}) \\ = \frac{1}{N}\sum_{n=1}^{N}\overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2}) \end{aligned} \tag{13}$$

We now turn to the final term on the right hand side of Equation (9):

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$

As we shall see, we will take advantage of the result from Lemma A.9 as well as the fact that $E\Big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\big|\boldsymbol{Z}\Big]=\mathbf{0}$. From Lemma A.9, we have:

$$\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)=$$
$$\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)-\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n)\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)\Big)+o_p(\mathrm{N}^{-1/2})\text{ for all }\boldsymbol{z}_n\in\boldsymbol{\mathcal{Z}}$$

where $\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)$ and $\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}_n)\mid\boldsymbol{z},\boldsymbol{\theta}\big)$ are defined in equations (♣ 4) and (♣ 5) in the proof of Lemma A.9. Since $\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)=\mathbf{0}$ for all $\boldsymbol{z}_n\notin\boldsymbol{\mathcal{Z}}$, we get:

$$\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)=$$
$$\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Big(\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)-\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n)\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\mid\boldsymbol{z}_n,\boldsymbol{\theta}_0\big)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$
$$+\,o_p(\mathrm{N}^{-1/2})\quad\text{for all }\boldsymbol{z}_n$$

From the definitions of $\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)$ and $\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z}_n)\mid\boldsymbol{z},\boldsymbol{\theta}\big)$, there exists $M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ such that we can express:

$$\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{z})\mid\boldsymbol{z},\boldsymbol{\theta}\big)=\frac{1}{Nh_N^L}\sum_{n=1}^{N}M(\boldsymbol{x}_n,\boldsymbol{z},\boldsymbol{\theta})K_h(\boldsymbol{z}_n-\boldsymbol{z})\quad\text{for all }\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}\text{ and }\boldsymbol{\theta}\in\boldsymbol{\Theta}$$

with $E\big[M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})\mid\boldsymbol{Z}\big]=\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z})\mid\boldsymbol{Z},\boldsymbol{\theta}\big)$ for all $\boldsymbol{Z}\in\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\theta}\in\boldsymbol{\Theta}$. Using assumptions (S1.3), (S2.2), (S2.3) and (S3.2) along with Lemma 4.1 we have that $M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ and $E\big[M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})\mid\boldsymbol{Z}\big]$ are bounded, M times differentiable functions of $\boldsymbol{Z}$ with bounded M derivatives everywhere in $\mathbb{S}(\boldsymbol{X})\times\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$ and $\boldsymbol{\mathcal{Z}}\times\boldsymbol{\Theta}$ respectively. Therefore, we can re-express:

$$\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)=$$
$$\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\left(\frac{1}{Nh_N^L}\sum_{m=1}^{N}\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n,\boldsymbol{\theta}_0)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n,\boldsymbol{\theta}_0)\mid\boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\right)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)$$
$$+\,o_p(\mathrm{N}^{-1/2})\quad\text{for all }\boldsymbol{z}_n$$

and consequently:

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}}^*_N(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)=$$

$$\frac{1}{N^2h_N^L}\sum_{n=1}^{N}\sum_{m\neq n}\bigg(\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n,\boldsymbol{\theta}_0)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n,\boldsymbol{\theta}_0)\mid\boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\bigg)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}$$

$$+\frac{K(0)}{Nh_N^L}\frac{1}{N}\sum_{n=1}^{N}\Big(M(\boldsymbol{x}_n,\boldsymbol{z}_n,\boldsymbol{\theta}_0)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n,\boldsymbol{\theta}_0)\mid\boldsymbol{z}_n\big]\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}+o_p(\mathrm{N}^{-1/2})$$

By the definition of our trimmed log-likelihood function $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\pi}\big)$, we know that we have $\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n,\boldsymbol{\theta}_0)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n,\boldsymbol{\theta}_0)\mid\boldsymbol{z}_n\big]\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)/f_{\boldsymbol{Z}}(\boldsymbol{z}_n)=\mathbf{0}$ whenever $\boldsymbol{z}_n\notin\boldsymbol{\mathcal{Z}}$. We have shown above that $\sup_n\Big\|\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\Big\|$ is finite w.p.1. Therefore, assumption (S3.2) and our trimming imply that $\sup_n\Big\|\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)/f_{\boldsymbol{Z}}(\boldsymbol{z}_n)\Big\|$ is also finite w.p.1. Combining this with the preceding discussion about $M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})$ and $E\big[M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta})\mid\boldsymbol{Z}\big]$ we know that there exists $\boldsymbol{C}>0$ such that:

$$\sup_n\left\|\Big(M(\boldsymbol{x}_n,\boldsymbol{z}_n,\boldsymbol{\theta}_0)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n,\boldsymbol{\theta}_0)\mid\boldsymbol{z}_n\big]\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\right\|<\boldsymbol{C}\quad\text{w.p.1}\tag{14}$$

which is sufficient for the expectation of this object to exist and be finite. To simplify notation, from now on we will denote $M(\boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}_0)\equiv M(\boldsymbol{X},\boldsymbol{Z})$. Then we have:

$$\frac{1}{N}\sum_{n=1}^{N}\Big(M(\boldsymbol{x}_n,\boldsymbol{z}_n)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid\boldsymbol{z}_n\big]\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}$$

$$\xrightarrow{p}\ E\left[\Big(M(\boldsymbol{X},\boldsymbol{Z})-E\big[M(\boldsymbol{X},\boldsymbol{Z})\mid\boldsymbol{Z}\big]\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{f_{\boldsymbol{Z}}(\boldsymbol{Z})}\right]=\mathbf{0}$$

where the last equality follows from the fact that $E\Big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\mid\boldsymbol{X},\boldsymbol{Z}\Big]=0$. Now, $K(0)$ is finite by assumption (S4.1), which combined with (S4.2.i) implies that $K(0)/(Nh_N^L)=o_p(\mathrm{N}^{-1/2})$ and therefore:

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}}^*_N(\boldsymbol{\theta}_0,\boldsymbol{z}_n)-\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)=$$

$$\frac{1}{N^2h_N^L}\sum_{n=1}^{N}\sum_{m\neq n}\bigg(\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n)-E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid\boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\bigg)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}$$

$$+o_p(\mathrm{N}^{-1/2})$$

We will now proceed to use a U-statistic representation. Let:

$$\begin{aligned}\widetilde{T}_N(\boldsymbol{w}_n,\boldsymbol{w}_m) &= \frac{1}{h_N^L}\Big(\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\\ &+ \frac{1}{h_N^L}\Big(\Big(M(\boldsymbol{x}_n,\boldsymbol{z}_m) - E\big[M(\boldsymbol{X},\boldsymbol{z}_m)\mid \boldsymbol{z}_m\big]\Big)K_h(\boldsymbol{z}_n-\boldsymbol{z}_m)\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_m,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_m)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)}\\ \text{and}\quad \widetilde{U}_N &= \binom{N}{2}^{-1}\sum_{1\le n<m\le N}\widetilde{T}_N(\boldsymbol{w}_n,\boldsymbol{w}_m)\end{aligned}$$

then, because $K(\cdot)$ is symmetric, we can re-express:

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\Big)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big) = \Big(\frac{N-1}{2N}\Big)\widetilde{U}_N + o_p(\mathrm{N}^{-1/2})$$

we will now determine $E\big[\widetilde{T}_N(\boldsymbol{w}_n,\boldsymbol{w}_m)\mid \boldsymbol{w}_n\big]$ starting with the first term on the right hand side of $\widetilde{T}_N(\boldsymbol{w}_n,\boldsymbol{w}_m)$. Using the iid nature of the data, we have:

$$\begin{aligned}&E\Bigg[\frac{1}{h_N^L}\Big(\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Bigg|\boldsymbol{w}_n\Bigg] =\\ &\Big(\int\Big(M(\boldsymbol{u},\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)\int K\big(\boldsymbol{\Psi}\big)f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}_n+h_N\boldsymbol{\Psi})d\boldsymbol{\Psi}d\boldsymbol{u}\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\end{aligned}$$

where $\boldsymbol{\Psi} = \big(\boldsymbol{v}-\boldsymbol{z}_n\big)/h_N$. We have shown previously that if assumptions (S2.2) and (S4) are satisfied, then

$$\int K\big(\boldsymbol{\Psi}\big)f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}_n+h_N\boldsymbol{\Psi})d\boldsymbol{\Psi} = f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}_n) + o_p(\mathrm{N}^{-1/2})\quad\text{for each } \boldsymbol{z}_n$$

therefore:

$$\begin{aligned}&E\Bigg[\frac{1}{h_N^L}\Big(\Big(M(\boldsymbol{x}_m,\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)K_h(\boldsymbol{z}_m-\boldsymbol{z}_n)\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}\Bigg|\boldsymbol{w}_n\Bigg]\\ &\quad= \Big(\int\Big(M(\boldsymbol{u},\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}_n)d\boldsymbol{u}\Big)'\frac{\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} + o_p(\mathrm{N}^{-1/2})\\ &\quad= o_p(\mathrm{N}^{-1/2})\end{aligned}\tag{15}$$

where the last equality follows from the fact that:

$$\int\Big(M(\boldsymbol{u},\boldsymbol{z}_n) - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big]\Big)\frac{f_{\boldsymbol{X},\boldsymbol{Z}}(\boldsymbol{u},\boldsymbol{z}_n)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)}d\boldsymbol{u} = E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big] - E\big[M(\boldsymbol{X},\boldsymbol{z}_n)\mid \boldsymbol{z}_n\big] = 0$$

we now turn to the second term on the right hand side of $\widetilde{T}_N(\boldsymbol{w}_n, \boldsymbol{w}_m)$. Using iterated expectations and the iid nature of the data we have:

$$E\left[\frac{1}{h_N^L}\left(\Big(M(\boldsymbol{x}_n, \boldsymbol{z}_m) - E\big[M(\boldsymbol{X}, \boldsymbol{z}_m) \mid \boldsymbol{z}_m\big]\Big) K_h(\boldsymbol{z}_n - \boldsymbol{z}_m)\right)' \frac{\nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_m, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_m)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)} \middle| \boldsymbol{w}_n\right] =$$

$$E\left[\frac{1}{h_N^L}\left(\Big(M(\boldsymbol{x}_n, \boldsymbol{z}_m) - E\big[M(\boldsymbol{X}, \boldsymbol{z}_m) \mid \boldsymbol{z}_m\big]\Big) K_h(\boldsymbol{z}_n - \boldsymbol{z}_m)\right)' \frac{E\Big[\nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_m, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_m)\big) \Big| \boldsymbol{x}_m, \boldsymbol{z}_m\Big]}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)} \middle| \boldsymbol{w}_n\right]$$

$$= E\left[\frac{1}{h_N^L}\left(\Big(M(\boldsymbol{x}_n, \boldsymbol{z}_m) - E\big[M(\boldsymbol{X}, \boldsymbol{z}_m) \mid \boldsymbol{z}_m\big]\Big) K_h(\boldsymbol{z}_n - \boldsymbol{z}_m)\right)' \mathbf{0} \middle| \boldsymbol{w}_n\right] = \mathbf{0} \tag{16}$$

where the last two equalities follow from the fact that $E\Big[\nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_m, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_m)\big) \Big| \boldsymbol{x}_m, \boldsymbol{z}_m\Big] = \mathbf{0}$ for all m. Combining equations (15-16) we have $E\big[\widetilde{T}_N(\boldsymbol{w}_n, \boldsymbol{w}_m)\big] = o_p(\mathrm{N}^{-1/2})$. Using equation (14) along with assumption (S4.1) -boundedness of $K(\cdot)$-, we can show that:

$$E\left\| \frac{1}{h_N^L}\left(\Big(M(\boldsymbol{x}_m, \boldsymbol{z}_n) - E\big[M(\boldsymbol{X}, \boldsymbol{z}_n) \mid \boldsymbol{z}_n\big]\Big) K_h(\boldsymbol{z}_m - \boldsymbol{z}_n)\right)' \frac{\nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_n)} \right.$$
$$\left. + \frac{1}{h_N^L}\left(\Big(M(\boldsymbol{x}_n, \boldsymbol{z}_m) - E\big[M(\boldsymbol{X}, \boldsymbol{z}_m) \mid \boldsymbol{z}_m\big]\Big) K_h(\boldsymbol{z}_n - \boldsymbol{z}_m)\right)' \frac{\nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_m, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_m)\big)}{f_{\boldsymbol{Z}}(\boldsymbol{z}_m)} \right\|^2 = \frac{1}{h_N^L} O(1)$$

or equivalently

$$E\big\|\widetilde{T}(\boldsymbol{w}_n, \boldsymbol{w}_m)\big\|^2 = \frac{1}{h_N^L} O(1) = o(N)$$

where the last equality follows from assumption (S4.2.i), which implies $N h_N^L \to \infty$. Therefore, the condition for the CLT of U-statistics is satisfied and we have $\widetilde{U}_N = o_p(\mathrm{N}^{-1/2})$. Combining equations (15-16) with this result, we get:

$$\frac{1}{N}\sum_{n=1}^{N}\Big(\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\Big)' \nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) = \left(\frac{N-1}{2N}\right)\widetilde{U}_N + o_p(\mathrm{N}^{-1/2}) \tag{17}$$
$$= o_p(\mathrm{N}^{-1/2})$$

We can finally go back to equation (5): combining equations (13) and (17), equation (9) becomes:

$$\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)' \nabla_{\boldsymbol{\pi}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)\Big] =$$
$$\frac{1}{N}\sum_{n=1}^{N} \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n) J_0(\boldsymbol{z}_n)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big) + o_p(\mathrm{N}^{-1/2})$$

and equation (5) becomes:

$$
\begin{aligned}
&-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}^*_N}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\\
=\quad&\frac{1}{N}\sum_{n=1}^{N}\Big[\nabla_{\boldsymbol{\theta}}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)\\
&\qquad+\overline{D}_{\mathcal{Z}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{x}_n,\boldsymbol{z}_n\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{z}_n\big]\Big)\Big]+o_p(\mathrm{N}^{-1/2})\\
=\quad&\frac{1}{N}\sum_{n=1}^{N}\Big[\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}}+\overline{D}_{\mathcal{Z}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{x}_n,\boldsymbol{z}_n\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{z}_n\big]\Big)\Big]+o_p(\mathrm{N}^{-1/2})
\end{aligned}
$$

By assumption (S6.1), $\boldsymbol{\theta}_0$ is in the interior of $\boldsymbol{\Theta}$. Therefore, Lemma 4.4 implies that $\boldsymbol{\theta}_0$ satisfies the first order conditions:

$$
E\Big[\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}\Big]=E\Big[\nabla_{\boldsymbol{\theta}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\Big]=\mathbf{0}
$$

using iterated expectations we also have:

$$
\begin{aligned}
&E\Big[\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)\Big]\\
&\qquad=E\Big[\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}E\Big[E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big|\boldsymbol{Z}\Big]\Big]=\mathbf{0}
\end{aligned}
$$

where the last equality follows from the fact that $E\Big[E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]\Big|\boldsymbol{Z}\Big]=E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]$. We also have:

$$
\begin{aligned}
&E\Big[\Big(\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}\Big)\Big(\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)\Big)'\Big]\\
&=E\Big[\Big(E\Big[\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}\Big|\boldsymbol{X},\boldsymbol{Z}\Big]\Big)\Big(\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)\Big)'\Big]\\
&=\mathbf{0}
\end{aligned}
$$

where the last equality follows from the fact that $E\Big[\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)/\partial\boldsymbol{\theta}\Big|\boldsymbol{X},\boldsymbol{Z}\Big]=\mathbf{0}$. We also showed previously that $\ell_{\mathcal{Z}}$ satisfies the information-identity result

$$
E\Big[\Big(\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}\Big)\Big(\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}\Big)'\Big]=-E\Big[\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big]=\Im_{\mathcal{Z}}
$$

Combining results shown previously, we know that

$$
E\Big\|\frac{\partial\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}}+\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)\Big\|^2
$$

exists and is finite. Therefore a Central Limit Theorem applies and we have:

$$\sqrt{N}\frac{1}{N}\sum_{n=1}^{N}\Big[\frac{\partial \ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}}+\overline{D}_{\mathcal{Z}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{x}_n,\boldsymbol{z}_n]-E\big[\boldsymbol{Y}\mid\boldsymbol{z}_n\big]\Big)\Big]\xrightarrow{d}\mathcal{N}\Big(\mathbf{0},\Im_{\mathcal{Z}}+\Omega\Big)$$

where

$$\Omega=$$

$$E\Big[\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}E\Big[\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)\Big(E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]-E\big[\boldsymbol{Y}\mid\boldsymbol{Z}\big]\Big)'\Big|\boldsymbol{Z}\Big]J_0\big(\boldsymbol{Z}\big)^{-1\prime}\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})'\Big]$$

$$=E\Big[\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})J_0\big(\boldsymbol{Z}\big)^{-1}\mathrm{Var}\Big[E[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}]\Big|\boldsymbol{Z}\Big]J_0\big(\boldsymbol{Z}\big)^{-1\prime}\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})'\Big]$$

Recall that $\overline{D}_{\mathcal{Z}}(\boldsymbol{Z})=E\Big[\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'}\Big|\ \boldsymbol{Z}\Big]$. Therefore

$$\Omega=$$

$$E\Bigg[E\Big[\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'}\Big|\ \boldsymbol{Z}\Big]J_0\big(\boldsymbol{Z}\big)^{-1}\mathrm{Var}\Big[E[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}]\Big|\boldsymbol{Z}\Big]J_0\big(\boldsymbol{Z}\big)^{-1\prime}E\Big[\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\pi}'}\Big|\ \boldsymbol{Z}\Big]'\Bigg]$$

From equation (6) we have

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\pi}_N^*}(\widetilde{\boldsymbol{\theta}},\boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\xrightarrow{p}-\Im_{\mathcal{Z}}$$

by (S6.2), $\Im_{\mathcal{Z}}$ is invertible. These results together imply that:

$$\sqrt{N}\big(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\big)\xrightarrow{d}\mathcal{N}\Big(\mathbf{0},\Im_{\mathcal{Z}}^{-1}\big(\Im_{\mathcal{Z}}+\Omega\big)\Im_{\mathcal{Z}}^{-1}\Big)=\mathcal{N}\Big(\mathbf{0},\Im_{\mathcal{Z}}^{-1}+\Im_{\mathcal{Z}}^{-1}\Omega\Im_{\mathcal{Z}}^{-1}\Big)\qquad(18)$$

which completes the proof of Theorem 1. □

Efficiency. Keeping the notation defined previously, let

$$\mathcal{D}_1=E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}\boldsymbol{\theta}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\Big]$$

$$\mathcal{D}_2(\boldsymbol{Z})=E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)+\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\ \Big|\ \boldsymbol{Z}\Big]$$

$$\mathcal{D}_3(\boldsymbol{Z})=-\nabla_{\boldsymbol{\theta}}\boldsymbol{\varphi}\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\mid\boldsymbol{Z},\boldsymbol{\theta}_0\big)$$

$$\mathcal{D}_4(\boldsymbol{Z})=J_0\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\mid\boldsymbol{Z},\boldsymbol{\theta}_0\big)$$

Note that the first object is an unconditional moment, while the remaining three are conditional ones. Define the matrices

$$\mathcal{D}(\boldsymbol{Z})=\begin{pmatrix}\mathcal{D}_1 & \mathcal{D}_2(\boldsymbol{Z})\\ \mathcal{D}_3(\boldsymbol{Z}) & \mathcal{D}_4(\boldsymbol{Z})\end{pmatrix}\quad,\quad\Sigma(\boldsymbol{Z})=\begin{pmatrix}\Im_{\mathcal{Z}} & \mathbf{0}\\ \mathbf{0} & \mathrm{Var}\Big[E\big[\boldsymbol{Y}\mid\boldsymbol{X},\boldsymbol{Z}\big]\ \Big|\ \boldsymbol{Z}\Big]\end{pmatrix}$$

Equation (22) in Ai and Chen (2003) follows the approach of Newey (1990) and shows the efficiency bound for models with conditional moment restrictions. Adapting their results, the efficiency bound for the model based on the moment conditions

$$E\left[\frac{\partial \ell_{\mathcal{Z}}(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z}))}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}$$

$$E\Big[\boldsymbol{\pi}^*(\boldsymbol{\theta},\boldsymbol{Z}) - E[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z},\boldsymbol{\theta}]\Big|\ \boldsymbol{Z}\Big] = \mathbf{0},$$

(a combination of unconditional and conditional moment restrictions) is given by the upper-left portion of $E\big[\mathcal{D}(\boldsymbol{Z})^{-1}\Sigma(\boldsymbol{Z})\mathcal{D}(\boldsymbol{Z})^{-1'}\big]$. From the proof of Lemma 4.1 we have $\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z}) = -\mathcal{D}_4(\boldsymbol{Z})^{-1}\mathcal{D}_3(\boldsymbol{Z})$. Using iterated expectations and the definition of the trimmed information matrix, we have $-\Im_{\mathcal{Z}} = E\big[\mathcal{D}_1 - \mathcal{D}_2(\boldsymbol{Z})\mathcal{D}_4(\boldsymbol{Z})^{-1}\mathcal{D}_3(\boldsymbol{Z})\big]$. Using these results and the formulas for the inverse of a partitioned matrix we get

$$\mathcal{D}(\boldsymbol{Z})^{-1} = \begin{pmatrix} -\Im_{\mathcal{Z}}^{-1} & \Im_{\mathcal{Z}}^{-1}\mathcal{B}_{\mathcal{Z}}(\boldsymbol{Z}) \\ \mathcal{A}_3(\boldsymbol{Z}) & \mathcal{A}_4(\boldsymbol{Z}) \end{pmatrix},$$

where $\mathcal{A}_3(\boldsymbol{Z})$ and $\mathcal{A}_4(\boldsymbol{Z})$ depend on $\mathcal{D}_1$, $\mathcal{D}_2(\boldsymbol{Z})$, $\mathcal{D}_3(\boldsymbol{Z})$ and $\mathcal{D}_4(\boldsymbol{Z})$. Consequently, the upper-left portion of $E\big[\mathcal{D}(\boldsymbol{Z})^{-1}\Sigma(\boldsymbol{Z})\mathcal{D}(\boldsymbol{Z})^{-1'}\big]$ is given by

$$\Im_{\mathcal{Z}}^{-1} + \Im_{\mathcal{Z}}^{-1}E\Big[\mathcal{B}_{\mathcal{Z}}(\boldsymbol{Z})\mathrm{Var}\Big[E\big[\boldsymbol{Y} \mid \boldsymbol{X},\boldsymbol{Z}\big]\ \Big|\ \boldsymbol{Z}\Big]\mathcal{B}_{\mathcal{Z}}(\boldsymbol{Z})'\Big]\Im_{\mathcal{Z}}^{-1},$$

which is precisely the asymptotic variance of $\widehat{\boldsymbol{\theta}}$.

Proof of Theorem 2

We proceed first by proving Lemma 4.3.

Proof of Lemma 4.3:

We have:

$$\widetilde{\pi}_{p_N}(\boldsymbol{z}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N}\frac{Y_{p_n}K_h(\boldsymbol{Z}_n - \boldsymbol{z})}{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})} \quad \text{for } p \in \{1,2\}$$

since $Y_{p_n} \in \{0,1\}$ for $p = 1,2$ then the conditions of Lemma A.1 are trivially met. Using A.2(A) and repeating steps parallel to those of parts (B)-(F) of such lemma we get that if assumptions (S2.1-2) and (S4) are satisfied then, since $E[\boldsymbol{Y}_p \mid \boldsymbol{Z}=\boldsymbol{z}] = \pi_p(\theta_0, \boldsymbol{z})$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ we get:

$$\underset{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}}{sup} \left|\widetilde{\pi}_{p_N}(\boldsymbol{z}) - \pi_p^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right| = o_p(\mathrm{N}^{-1/4}) \quad \text{for } p \in \{1,2\}$$

We have $\overline{\pi}_{p_N}(\boldsymbol{z}) = \text{Max}\left\{0, \text{Min}\left\{\widetilde{\pi}_{p_N}(\boldsymbol{z}), 1\right\}\right\}$ and $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \equiv \left(\overline{\pi}_{1_N}(\boldsymbol{z}), \overline{\pi}_{2_N}(\boldsymbol{z})\right)'$. By Lemma 4.1, there exists $0 < b < 1$ such that for each $p \in \{1,2\}$: $b < \pi_p^*(\boldsymbol{\theta}_0, \boldsymbol{z}) < 1 - b$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. In other words, for each $p \in \{1,2\}$: $\inf\limits_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left(\pi_p^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right) > b$ and $1 - \sup\limits_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \pi_p^*(\boldsymbol{\theta}_0, \boldsymbol{z}) > b$. Consequently:

$$\underset{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}}{sup} \left\|\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right\| = o_p(\mathrm{N}^{-1/4}) \tag{19}$$

where $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \equiv \left(\overline{\pi}_{1_N}(\boldsymbol{z}), \overline{\pi}_{2_N}(\boldsymbol{z})\right)'$. By definition, $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}$. Therefore, it shares two very important properties with $\widehat{\boldsymbol{\pi}}_N(\boldsymbol{\theta}, \boldsymbol{z})$: they both converge uniformly in probability at the sate rate, and all the intermediate values between $\overline{\pi}_N(\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})$ are in $[0,1]^2$. This allows us to take advantage of assumption (S3.2) -which holds for $\boldsymbol{\pi} \in [0,1]^2$- which yields uniqueness of equilibrium and establishes a uniform bound for $\left\|J\left(\boldsymbol{\pi}^*(\boldsymbol{\theta}, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\right\|$ in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Take $(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. A mean-value approximation yields:

$$\widehat{J}_N\left(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1} = \widehat{J}_N\left(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1} + \Big(\boldsymbol{I}_2 \otimes \left(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right)\Big)'\Big(\nabla_{\boldsymbol{\pi}} \widehat{J}_N\left(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\Big)$$

where $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z})$ is between $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})$, the matrix $\boldsymbol{I}_2$ is a 2×2 identity matrix and -following the rules of matrix differentiation- $\nabla_{\boldsymbol{\pi}} \widehat{J}_N\left(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1} = \nabla_{\boldsymbol{\pi}} \text{vec}\Big(\widehat{J}_N\left(\boldsymbol{\pi} \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\Big)$. Using Lemma A.2 along with assumption (S3.2) and the fact that $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, we can show that $\sup\limits_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\nabla_{\boldsymbol{\pi}} \widehat{J}_N\left(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\right\| = O_p(1)$. Combining this with equation (19), we get:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\Big(\boldsymbol{I}_2 \otimes \left(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right)\Big)'\Big(\nabla_{\boldsymbol{\pi}} \widehat{J}_N\left(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\Big)\right\| = o_p(\mathrm{N}^{-1/4})$$

using Lemma A.4 and the fact that $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, we have

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{J}_N\left(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1} - J\left(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\right\| = o_p(\mathrm{N}^{-1/4})$$

Then, using the mean-value approximation we get:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{J}_N\left(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1} - J\left(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\right)^{-1}\right\| = o_p(\mathrm{N}^{-1/4}) \tag{20}$$

we next examine the term $\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$. A mean-value approximation yields:

$$\big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\big] = \big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\big] - \widehat{J}_N\big(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big)$$

where (once again) $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z})$ is between $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$. We have $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \in [0,1]^2$ and $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Since $[0,1]^2$ is a compact set, Lemma A.2 yields:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widehat{J}_N\big(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\| = O_p(1) \quad \text{and} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big\| = o_p(\mathrm{N}^{-1/4})$$

Combining these results with (19), we get:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big] - \Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]\Big\| = o_p(\mathrm{N}^{-1/4}) \tag{21}$$

Combining (20) and (21) we have:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big] - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]\Big\| = o_p(\mathrm{N}^{-1/4}) \tag{22}$$

We defined:

$$\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) = \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big]$$

$$\rho(\boldsymbol{\theta},\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]$$

Therefore, (19) and (22) yield:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \rho(\boldsymbol{\theta},\boldsymbol{z})\Big\| = o_p(\mathrm{N}^{-1/4})$$

which proves part (A) of Lemma 4.3. To prove part (B) of the lemma, we first show that if assumptions (S1.3), (S2), (S3) and (S4) are satisfied, then all the results of Lemma A.7 hold replacing $\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{z},\boldsymbol{\theta})$ with $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta})$ with $\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0)$. That is:

$$\text{(A)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big|\widehat{\varphi}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \varphi_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\Big| = o_p(\mathrm{N}^{-1/4})$$

$$\text{(B)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big|\widehat{\delta}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta_p\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\Big| = o_p(\mathrm{N}^{-1/4})$$

$$
\begin{aligned}
&\text{(C)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{\delta}^{(m)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \delta^{(m)}_{p}\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right| && = o_p(\mathrm{N}^{-1/4}) \quad m = 1,\ldots,M\\
&\text{(D)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\zeta}^{(m)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \zeta^{(m)}_{p}\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right\| && = o_p(\mathrm{N}^{-1/4}) \quad m = 0,\ldots,M \qquad (23)\\
&\text{(E)} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\|\widehat{\xi}^{(m)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big) - \xi^{(m)}_{p}\big(\pi^*_{\text{-}p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)\right\| && = o_p(\mathrm{N}^{-1/4}) \quad m = 0,\ldots,M
\end{aligned}
$$

where each of these objects was defined in Lemma A.2 and the paragraph immediately preceding it. The details of the proof are completely parallel to those used in the proof of Lemma A.7. We proceed by taking mean-value approximations and take advantage of two key results: the uniform rate of convergence in (19) as well as the fact that $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}$, which in turn implies that all the mean-values are also in $[0,1]^2$. Just like in the proof of Lemma A.7, compactness of $[0,1]^2 \times \boldsymbol{\mathcal{Z}} \times \mathbb{S}(\boldsymbol{X})$ allows us to use Lemma A.2 and obtain results (A)-(E). Now, following our notation, let $\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)$ be the determinant of $\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)$. Then by equation (23 (B)) we have $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - d\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\right| = o_p(\mathrm{N}^{-1/4})$. Next, we proceed as in the proof of Lemma A.8 by noting that if assumption (S3.2) is satisfied, then $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{d}_N\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\right| = O_p(1)$, $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1} - d\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\right| = o_p(\mathrm{N}^{-1/4})$ for all $\boldsymbol{\pi} \in [0,1]^2$. Since $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}$, we get $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\right| = O_p(1)$ and $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left|\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1} - d\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\right| = o_p(\mathrm{N}^{-1/4})$. Next, notice that

$$
\begin{aligned}
\underset{2\times(k+2)}{\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})} &= \Big(\boldsymbol{I}_2 \otimes \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big]\Big)' \nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)\\
&\quad + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\\
\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})} &= \Big(\boldsymbol{I}_2 \otimes \nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)' \nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)\\
&\quad + \Big(\boldsymbol{I}_{2(k+2)} \otimes \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big]\Big)' \nabla_{\boldsymbol{\theta}}\mathrm{vec}\,\Big\{\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)\Big\}\\
&\quad + \nabla_{\boldsymbol{\theta}}\mathrm{vec}\,\Big(\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)
\end{aligned}
$$

and

$$\underset{2\times(k+2)}{\nabla_{\boldsymbol{\theta}}\rho(\boldsymbol{\theta},\boldsymbol{z})} = \Big(\boldsymbol{I}_2 \otimes \Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0)\Big]\Big)' \nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\ \Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)$$

$$+ J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)$$

$$\underset{2(k+2)\times(k+2)}{\nabla_{\boldsymbol{\theta\theta'}}\rho(\boldsymbol{\theta},\boldsymbol{z})} = \Big(\boldsymbol{I}_2 \otimes \nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)' \nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\ \Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)$$

$$+ \Big(\boldsymbol{I}_{2(k+2)} \otimes \Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0)\Big]\Big)' \nabla_{\boldsymbol{\theta}}\mathrm{vec}\ \Big\{\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\ \Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big)\Big\}$$

$$+ \nabla_{\boldsymbol{\theta}}\mathrm{vec}\ \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)\Big)$$

From assumptions (S1.3) - $G_1(\cdot)$, $G_2(\cdot)$ bounded functions with bounded $M+2$ derivatives-, (S2.3) - supports $\mathbb{S}(\boldsymbol{X}_1)$ and $\mathbb{S}(\boldsymbol{X}_2)$ being compact sets- and (S3.2) - $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big\|J(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})^{-1}\Big\|$ being finite for all $\boldsymbol{\pi} \in [0,1]^2$-, we have:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big\|\nabla_{\boldsymbol{\theta}}\rho(\boldsymbol{\theta},\boldsymbol{z})\Big\| < \widetilde{C}_1 \quad \text{and} \quad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big\|\nabla_{\boldsymbol{\theta\theta'}}\rho(\boldsymbol{\theta},\boldsymbol{z})\Big\| < \widetilde{C}_2 \quad \text{w.p.1}$$

for some constants $\widetilde{C}_1 > 0$, $\widetilde{C}_2 > 0$.

Now, we have that $\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ and $\nabla_{\boldsymbol{\theta\theta'}}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})$ depend exclusively on the terms:

$$\widehat{d}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}, \quad \widehat{\delta}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \overline{\pi}_{\text{-}p_N}(\boldsymbol{z})\widehat{\zeta}^{(0)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

$$\widehat{\xi}^{(1)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \widehat{\zeta}^{(1)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \overline{\pi}_{\text{-}p_N}(\boldsymbol{z})\widehat{\zeta}^{(1)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

$$\overline{\pi}_{\text{-}p_N}(\boldsymbol{z})^2\widehat{\delta}^{(1)}_{p_N}\big(\overline{\pi}_{\text{-}p_N}(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

for $p \in \{1,2\}$.

While $\nabla_{\boldsymbol{\theta}}\rho(\boldsymbol{\theta},\boldsymbol{z})$ and $\nabla_{\boldsymbol{\theta\theta'}}\rho(\boldsymbol{\theta},\boldsymbol{z})$ depend on the exact same way on the terms:

$$d\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}, \quad \delta_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0)\zeta^{(0)}_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

$$\xi^{(1)}_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \zeta^{(1)}_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big), \quad \boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0)\zeta^{(1)}_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

$$\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0)^2\delta^{(1)}_p\big(\boldsymbol{\pi}^*_{-p}(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_p\big)$$

for $p \in \{1,2\}$. Therefore, using (19), (23) and part (A) of Lemma 4.3 -shown above- we get:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta}} \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z}) \right\| = o_p(\mathrm{N}^{-1/4})$$

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta\theta'}} \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) - \nabla_{\boldsymbol{\theta\theta'}} \boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z}) \right\| = o_p(\mathrm{N}^{-1/4})$$

which establishes part (B) of the lemma. To show part (C), note that by definition of the equilibrium conditions:

$$\left[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \right] = \mathbf{0} \quad \text{for all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$$

and consequently:

$$\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \quad \text{for all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$$

$$\nabla_{\boldsymbol{\theta}} \rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1} \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) = \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \quad \text{for all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$$

This completes the proof of Lemma 4.3.

Note that:

$$\nabla_{\boldsymbol{\theta\theta'}} \rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \nabla_{\boldsymbol{\theta\theta'}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + \Big(\boldsymbol{I}_2 \otimes \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) \Big)' \nabla_{\boldsymbol{\theta}} \Big(\mathrm{vec} \left\{ J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1} \right\} \Big)$$

since $\nabla_{\boldsymbol{\theta\theta'}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = \nabla_{\boldsymbol{\theta}} \mathrm{vec} \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1} \nabla_{\boldsymbol{\theta}} \varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) \Big)$. As we will see below, the fact that $\nabla_{\boldsymbol{\theta\theta'}} \rho(\boldsymbol{\theta}_0,\boldsymbol{z}) \neq \nabla_{\boldsymbol{\theta\theta'}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$ will not affect our asymptotic results since $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \left\| \nabla_{\boldsymbol{\theta\theta'}} \rho(\boldsymbol{\theta},\boldsymbol{z}) \right\| < \widetilde{C}_2$ w.p.1 for some $\widetilde{C}_2 > 0$ implies that the only term in which $\nabla_{\boldsymbol{\theta\theta'}} \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z})$ shows up in the proof of Theorem 2 goes to zero in probability at the appropriate rate.

The next step is to show that $\widetilde{\boldsymbol{\pi}}_N(\boldsymbol{\theta},\boldsymbol{z})$ satisfies the result of Lemma A.6 when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

Lemma A.10 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Take $(\boldsymbol{\theta},\boldsymbol{z}) \in \boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ and let*

$$\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) = \underset{\boldsymbol{\pi}\in[0,1]^2}{\mathrm{argmax}}\ \widehat{\boldsymbol{Q}}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta})$$

$$\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) = \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1} \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \Big]$$

Then $\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}) \right\| = o_p(N^{-1/2})$ *and it follows from Lemma A.6 that:*

(A) $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big] + o_p(N^{-1/2})$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$.

(B) As in the proof of Lemma A.2(B), define:

$$\widehat{S}_{p_N}(\pi_{-p} \mid \boldsymbol{z},\boldsymbol{\theta}_p) = \frac{1}{Nh_N^L}\sum_{n=1}^{N} G_p\big(\boldsymbol{X}'_{pn}\boldsymbol{\beta}_p + \alpha_p\pi_{-p}\big)K_h(\boldsymbol{Z}_n - \boldsymbol{z}) \quad \text{for } p \in \{1,2\}$$

and let $\widehat{S}_N(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}) = \big(\ \widehat{S}_{1_N}(\boldsymbol{\pi}_2 \mid \boldsymbol{z},\boldsymbol{\theta}_1)\ ,\ \widehat{S}_{2_N}(\boldsymbol{\pi}_1 \mid \boldsymbol{z},\boldsymbol{\theta}_2)\ \big)'$. *Then:*

$$\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big] + o_p(N^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$.

Proof: A second-order approximation yields:[31]

$$\begin{aligned}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) &= \widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big] \\ &- \frac{1}{2}\bigg(\boldsymbol{I}_2 \otimes \Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]\bigg)'\nabla_{\boldsymbol{\pi}}\text{vec}\ \Big[J\big(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big]\bigg(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\bigg)\end{aligned}$$

with $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z})$ between $\overline{\boldsymbol{\pi}}_N(\boldsymbol{z})$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$. As we argued previously, (S1.3), (S2.3) and (S3.2) imply that $\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} \Big\|\nabla_{\boldsymbol{\pi}}\text{vec}\ \Big[J\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big]\Big\| = O_p(1)$ for all $\boldsymbol{\pi} \in [0,1]^2$. Therefore, since we have $\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ this implies: $\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \Big\|\nabla_{\boldsymbol{\pi}}\text{vec}\ \Big[J\big(\widetilde{\overline{\boldsymbol{\pi}}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big]\Big\| = O_p(1)$. Combining this with eq. (19), the second order approximation yields:

$$\begin{aligned}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) =& \widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big] \\ &+ o_p(\text{N}^{-1/2})\end{aligned}$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Using this result and the fact that $\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \Big\|\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\| = O_p(1)$ (from eq. (20)) we have:

$$\begin{aligned}&\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big] = \\ &\qquad\bigg(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1} + \Big[\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big]\bigg) \\ &\qquad\quad \times \bigg(\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]\bigg) + o_p(\text{N}^{-1/2})\end{aligned}$$

[31] Recall that by definition, $J\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) = \nabla_{\boldsymbol{\pi}}\Big(\boldsymbol{\pi} - \varphi\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big)$.

We have $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \in [0,1]^2$ -a compact set- for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Therefore, using Lemma A.2 and the equilibrium conditions ($\boldsymbol{\varphi}\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)$ for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$) we have:

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| \widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \right\| = o_p(\mathrm{N}^{-1/4})$$

-we had already shown this result holds uniformly in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$ in the proof of Lemma A.6-. Combining this with eq. (20) we get:

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| \Big(\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big) \Big(\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big) \right\| = o_p(\mathrm{N}^{-1/4})$$

Using equations (19)-(20) and the fact that $\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) \right\| = O_p(1)$, we also have:

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| \Big(\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} - J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big) \Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) \Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] \Big) \right\| = o_p(\mathrm{N}^{-1/4})$$

Therefore:

$$\begin{aligned} \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \Big] &= J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] \\ &\quad - \Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Therefore:

$$\begin{aligned} \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) &= \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big)^{-1} \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \Big] \\ &= \overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) + J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] \\ &\quad - \Big[\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Simplifying the last expression yields:

$$\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. From the proof of Lemma A.6 we have:

$$\widehat{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1} \Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \Big] + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. This proves part (A) of the lemma. Part (B) follows immediately from the proof of Lemma A.6. □

Before proceeding, we next show that $\nabla_{\boldsymbol{\theta}} \widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0, \boldsymbol{z})$ satisfies a result analogous to that of Lemma A.9:

Lemma A.11 *Let $\boldsymbol{\mathcal{Z}}$ be as defined in (S3.2) and suppose assumptions (S1.3), (S2), (S3) and (S4) are satisfied. Then there exist matrices* $\underset{2\times(k+2)}{\widehat{V}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)}$ *and* $\underset{2\times(k+2)}{\Phi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)}$ *such that*

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \left\| \widehat{V}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - f_{\boldsymbol{Z}}(\boldsymbol{z})\Phi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) \right\| = o_p(N^{-1/4})$$

and

$$\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta}_0,\boldsymbol{z}) - \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{V}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\Phi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big] + o_p(N^{-1/2})$$

for all $\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}$.

Proof:

Recall that

$$\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}_N^*}(\boldsymbol{\theta},\boldsymbol{z}) = \Big(\boldsymbol{I}_2 \otimes \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big]\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\Big\}\Big) + \widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)$$

Using Lemma A.2, equation (19), assumptions (S1.3), (S2.3), (S3.2) and the fact that $\overline{\boldsymbol{\pi}}(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}$, we can take the same steps as those of the proof of Lemma A.9 to show that:

$$\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) = J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) + \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big] + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}$, with $\widehat{W}_N\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}\big)$ and $\Gamma\big(\boldsymbol{\pi} \mid \boldsymbol{z},\boldsymbol{\theta}\big)$ exactly as defined in the proof of Lemma A.9.

Once again, using Lemma A.2, equation (19), assumptions (S1.3), (S2.3), (S3.2) and the fact that $\overline{\boldsymbol{\pi}}(\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}$ and $\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \in [0,1]^2$ for all $\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}$, we can show that:

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}}\left\|\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\}\Big) - \nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\}\Big)\right\| = o_p(\mathrm{N}^{-1/4})$$

which combined with equation (21) and the equilibrium condition $\Big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big] = \mathbf{0}$ for all $\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}$, we get:

$$\Big(\boldsymbol{I}_2 \otimes \Big[\widehat{\varphi}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \overline{\boldsymbol{\pi}}_N(\boldsymbol{z})\Big]\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{\widehat{J}_N\big(\overline{\boldsymbol{\pi}}_N(\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\}\Big)$$
$$= \Big(\boldsymbol{I}_2 \otimes \Big[\widehat{\varphi}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\Big]\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\}\Big) + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. From the second-to-last equation in the proof of Lemma A.6 we have:

$$\widehat{\varphi}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) = \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\left[\widehat{S}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right] + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$.

Combining these results, we have:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}}\widehat{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0, \boldsymbol{z}) &= J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) \\
&+ \frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\Big[\widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)\Big] \\
&+ \Big(\boldsymbol{I}_2 \otimes \Big[\frac{1}{f_{\boldsymbol{Z}}(\boldsymbol{z})}\left[\widehat{S}_N(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0) - \widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z})\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\right]\Big]\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}\Big\}\Big) + o_p(\mathrm{N}^{-1/2})
\end{aligned}$$

The proof is complete by noting that $J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) = \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})$ and letting:

$$\begin{aligned}
\widehat{V}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) &= \widehat{W}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) \\
&\quad + \Big(\boldsymbol{I}_2 \otimes \widehat{S}_N\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}\Big\}\Big) \\
\Phi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) &= \Gamma\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big) \\
&\quad + \Big(\boldsymbol{I}_2 \otimes \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z})\Big)'\nabla_{\boldsymbol{\theta}}\Big(\mathrm{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}\Big\}\Big)
\end{aligned}$$

□

We are now ready to prove Theorem 2.

Proof of Theorem 2:

From Lemma 4.1 and assumption (S3.2), $\rho(\boldsymbol{\theta}, \boldsymbol{Z})$ is continuous in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$. Combining this with the continuity of the linear function $\boldsymbol{X}'\boldsymbol{\beta} + \alpha\pi$ and assumption (S1.3), then $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ is continuous in $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. By assumptions (S2.3) and (S3), the set $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$ is compact and therefore the continuity of $\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ is uniform in $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. In addition, using once again Lemma 4.1 and assumption (S3.2), the compactness of $\boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$ implies that there exists a $\boldsymbol{C} > 0$ such that $\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \Big\|\boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{z})\Big\| < \boldsymbol{C}$ w.p.1. Let $\boldsymbol{\mathcal{C}} = \big\{\boldsymbol{v} \in \mathbb{R}^2 : \|\boldsymbol{v}\| \leq \boldsymbol{C}\big\}$. Now, take any $\boldsymbol{w} \in \big\{0, 1\big\} \times \mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$ and any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ with the corresponding $\boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{z}) \in \boldsymbol{\mathcal{C}}$. Then, by uniform continuity we have that for all $M > 0$ there exists $\delta > 0$ such that $\boldsymbol{\rho} \in \boldsymbol{\mathcal{C}}$ and $\Big\|\boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{z}) - \boldsymbol{\rho}\Big\| < \delta$

imply $\left\|\ell_{\mathcal{Z}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\big)-\ell_{\mathcal{Z}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\rho}\big)\right\|<M$. Now let $\widetilde{\delta}=\min\left\{\delta\ ,\ \boldsymbol{C}-\sup_{\substack{\boldsymbol{z}\in\mathcal{Z}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\|\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\|\right\}$. Then $\widetilde{\delta}>0$ and using Lemma 4.3(A) we have that for all $\varepsilon>0$, there exists $N_{\widetilde{\delta}}$ such that $N>N_{\widetilde{\delta}}$ implies:

$$\Pr\left\{\sup_{\substack{\boldsymbol{z}\in\mathcal{Z}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\left\|\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})-\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\right\|>\widetilde{\delta}\right\}<\varepsilon$$

Therefore, $N>N_{\widetilde{\delta}}$ implies

$$\Pr\left\{\sup_{\substack{\boldsymbol{w}\in\{0,1\}\times\mathbb{S}(\boldsymbol{X})\times\mathcal{Z}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\left\|\ell_{\mathcal{Z}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\big)-\ell_{\mathcal{Z}}\big(\boldsymbol{w},\boldsymbol{\theta},\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z})\big)\right\|\geq M\right\}<\varepsilon$$

and consequently:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)-\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)\right|\xrightarrow{p}0$$

From assumption (S4.3), the sample is iid. As we mentioned above, Lemma 4.1, assumption (S3.2) and the continuity of the linear function $\boldsymbol{\beta}'\boldsymbol{X}+\alpha\pi$, imply that $\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)$ is a continuous function at each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ with probability one. By (S3.1), $\boldsymbol{\Theta}$ is compact. We also know that $\rho(\boldsymbol{\theta},\boldsymbol{Z})\in\boldsymbol{\mathcal{C}}$ (a compact set) for all $\boldsymbol{\theta}\in\boldsymbol{\Theta}$ and all $\boldsymbol{Z}\in\mathcal{Z}$. Compactness of $\{0,1\}\times\mathbb{S}(\boldsymbol{X})\times\mathcal{Z}\times\boldsymbol{\mathcal{C}}$ implies that there exists $\bar{\ell}$ such that $\left|\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)\right|<\bar{\ell}$ with probability one. These properties are sufficient to satisfy the assumptions of Lemma 2.4 in Newey and McFadden (1994) (dominated uniform convergence theorem) and imply that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{z}_n)\big)-E\big[\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]\right|=o_p(1)$$

These results together imply that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widetilde{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)-E\big[\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]\right|=o_p(1)$$

From Lemma 4.5 we know that $E\big[\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]$ is uniquely maximized at $\boldsymbol{\theta}_0$. By Lemma 4.1 and assumption (S3.2), we know that $E\big[\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\rho(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\big]$ is continuous. The result immediately above showed that $\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widetilde{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)$ converges in probability to $E\big[\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\rho(\boldsymbol{\theta},\boldsymbol{Z})\big)\big]$ uniformly in $\boldsymbol{\Theta}$. Since $\widetilde{\boldsymbol{\theta}}$ maximizes $\frac{1}{N}\sum_{n=1}^{N}\ell_{\mathcal{Z}}\big(\boldsymbol{w}_n,\boldsymbol{\theta},\widetilde{\boldsymbol{\pi}^*}(\boldsymbol{\theta},\boldsymbol{z}_n)\big)$ in $\boldsymbol{\Theta}$, all the conditions of Theorem 2.1 in Newey and McFadden are met and therefore $\widetilde{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}_0$.

Proof of Theorem 2(B):

With probability approaching one uniformly in $\boldsymbol{\Theta} \times \boldsymbol{\mathcal{Z}}$, the estimator $\widetilde{\boldsymbol{\theta}}$ satisfies the first order conditions:

$$\frac{1}{N}\sum_{n=1}^{N}\Big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)\Big\} = \mathbf{0}$$

and (using Lemma 4.1, along with assumption (S3.2)), $\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z})$ is an M times differentiable function of $\boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and for all $\boldsymbol{z}_n$ (since $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$ for all $\boldsymbol{z}_n$). A first order Taylor series approximation for $\widetilde{\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_0$ yields:

$$\begin{aligned} &-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widehat{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \\ &\qquad \frac{1}{N}\sum_{n=1}^{N}\Big\{\nabla_{\boldsymbol{\theta}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)'\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)\Big\} \end{aligned} \tag{24}$$

with $\overline{\boldsymbol{\theta}}$ between $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ and:

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \\ &\frac{1}{N}\sum_{n=1}^{N}\Bigg[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) + \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n) \\ &+ \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\widetilde{\boldsymbol{\pi}^*_N}(\boldsymbol{\theta}, \boldsymbol{z}_n)'\big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \otimes \boldsymbol{I}_{(k+2)}\big] + \nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)'\Big\{\nabla_{\boldsymbol{\pi}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \\ &+ \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)\nabla_{\boldsymbol{\theta}}\widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\Big\}\Bigg] \end{aligned}$$

where $\boldsymbol{I}_{(k+2)}$ is a $(k+2) \times (k+2)$ identity matrix.

We have:

$$\begin{aligned} &\left\|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - E\left[\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right\| \\ &\qquad \leq \sup_n \left\|\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}^*_N}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| \\ &\qquad\quad + \left\|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - E\left[\frac{\partial^2\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right\| \end{aligned}$$

Lemma 4.1(A), assumptions (S1.3), (S3.2) and the compactness of $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$ imply that the functions $\nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$, $\nabla_{\boldsymbol{\theta\theta'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$, $\nabla_{\boldsymbol{\theta\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ and $\nabla_{\boldsymbol{\pi\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{Z})\big)$ are all uniformly continuous in $\mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}} \times \boldsymbol{\Theta}$. Since $\overline{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$ then using Lemma 4.2(A) and taking the same steps as above we get: $\sup_n \Big\| \nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) - \nabla_{\boldsymbol{\theta}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \Big\| = o_p(1)$, $\sup_n \Big\| \nabla_{\boldsymbol{\theta\theta'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) - \nabla_{\boldsymbol{\theta\theta'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \Big\| = o_p(1)$, $\sup_n \Big\| \nabla_{\boldsymbol{\theta\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) - \nabla_{\boldsymbol{\theta\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \Big\| = o_p(1)$, and $\sup_n \Big\| \nabla_{\boldsymbol{\pi\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) - \nabla_{\boldsymbol{\pi\pi'}} \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big) \Big\| = o_p(1)$.

The results in Lemma 4.3(B) and the trimming index $\mathbb{1}\{\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}\}$ imply that $\sup_n \Big\| \nabla_{\boldsymbol{\theta}} \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n) \Big\| = o_p(1)$ and $\sup_n \Big\| \nabla_{\boldsymbol{\theta\theta'}} \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n) - \nabla_{\boldsymbol{\theta\theta'}} \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n) \Big\| = o_p(1)$. These results together imply:

$$\sup_n \left\| \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| = o_p(1)$$

$\overline{\boldsymbol{\theta}}$ is intermediate between $\widetilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Therefore $\overline{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$. As we argued in the proof of part (A) of the theorem (a few paragraphs above), from Lemma 4.1 and assumption (S3.2) we know that $\sup_{\substack{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}} \\ \boldsymbol{\theta} \in \boldsymbol{\Theta}}} \big\| \boldsymbol{\rho}(\widetilde{\boldsymbol{\theta}}, \boldsymbol{z}) \big\| < \boldsymbol{C}$ for some $\boldsymbol{C} > 0$. Combining this with assumptions (S1.3), (S2.3) we know that $\Big\| \partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{\theta}, \boldsymbol{z}_n)\big) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' \Big\|$ is bounded with probability one for all $\boldsymbol{w}_n$, all $\boldsymbol{z}_n \in \boldsymbol{\mathcal{Z}}$ and all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By Lemma 4.1 and assumption (S3.2) it is also a continuous function everywhere in $\boldsymbol{\Theta}$. Consequently, $E\Big[\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'\Big]$ is continuous and bounded. Once again using Lemma 2.4 in Newey and McFadden, we get:

$$\left\| \frac{1}{N} \sum_{n=1}^{N} \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \boldsymbol{\rho}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - E\left[\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \right\| \xrightarrow{p} 0$$

and consequently:

$$\frac{1}{N} \sum_{n=1}^{N} \frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xrightarrow{p} E\left[\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \tag{25}$$

We have:

$$\frac{\partial^2 \ell_{\mathcal{Z}}\big(\boldsymbol{w},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\big)\nabla_{\boldsymbol{\theta}}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})$$

$$+ \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})'\big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\big)\otimes\boldsymbol{I}_{(k+2)}\big] + \nabla_{\boldsymbol{\theta}}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})'\Big\{\nabla_{\boldsymbol{\pi}\boldsymbol{\theta}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\big)$$

$$+ \nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta},\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\big)\nabla_{\boldsymbol{\theta}}\boldsymbol{\rho}(\boldsymbol{\theta},\boldsymbol{Z})\Big\}\Big]$$

Now recall that

$$\rho(\boldsymbol{\theta},\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big)^{-1}\big[\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})\big]$$

From the equilibrium conditions, we have: $\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \mid \boldsymbol{z},\boldsymbol{\theta}\big) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = \mathbf{0}$ for all $\boldsymbol{z}\in\mathcal{Z}$. As we pointed out in the proof of Lemma 4.3, this yields:

$$\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \quad \text{for all } \boldsymbol{z}\in\mathcal{Z}$$

$$\nabla_{\boldsymbol{\theta}}\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big) = \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) \quad \text{for all } \boldsymbol{z}\in\mathcal{Z}$$

We do not have $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z})$ for all $\boldsymbol{z}\in\mathcal{Z}$, for:
$\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\rho(\boldsymbol{\theta}_0,\boldsymbol{z}) = \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) + \Big(\boldsymbol{I}_2\otimes\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big)'\nabla_{\boldsymbol{\theta}}\Big(\text{vec}\,\Big\{J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\Big\}\Big)$
since $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{z}) = \nabla_{\boldsymbol{\theta}}\text{vec}\,\Big(J\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)^{-1}\nabla_{\boldsymbol{\theta}}\varphi\big(\boldsymbol{\pi}^*(\boldsymbol{z},\boldsymbol{\theta}_0) \mid \boldsymbol{z},\boldsymbol{\theta}_0\big)\Big)$. However, from Lemma 4.2 and assumptions (S1.3), (S2.3) and (S3.2) we have: $\sup_{\substack{\boldsymbol{z}\in\mathcal{Z}\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}}\Big\|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\rho(\boldsymbol{\theta},\boldsymbol{z})\Big\| < \boldsymbol{D}$ for some $\boldsymbol{D} > 0$. This is sufficient for $E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})'\big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\otimes\boldsymbol{I}_{(k+2)}\big]\Big]$ to exist. Using iterated expectations we have:

$$E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})'\big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\otimes\boldsymbol{I}_{(k+2)}\big]\Big] =$$
$$E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})'\Big[E\Big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\otimes\boldsymbol{I}_{(k+2)}\Big] \mid \boldsymbol{X},\boldsymbol{Z}\Big]\Big] =$$
$$E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\rho(\boldsymbol{\theta}_0,\boldsymbol{Z})'\Big[E\Big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big)\otimes\boldsymbol{I}_{(k+2)}\Big] \mid \boldsymbol{X},\boldsymbol{Z}\Big]\Big] = \mathbf{0}$$

where the second-to-last equality uses the fact (mentioned above) that $\boldsymbol{\rho}(\boldsymbol{\theta}_0,\boldsymbol{Z}) = \boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})$ everywhere in $\mathcal{Z}$ and the last equality uses the fact that $E\Big[\nabla_{\boldsymbol{\pi}}\ell_{\mathcal{Z}}\big(\boldsymbol{W},\boldsymbol{\theta}_0,\boldsymbol{\pi}^*(\boldsymbol{\theta}_0,\boldsymbol{Z})\big) \mid \boldsymbol{X},\boldsymbol{Z}\Big] = \mathbf{0}$

for all $(\boldsymbol{X}, \boldsymbol{Z}) \in \mathbb{S}(\boldsymbol{X}) \times \boldsymbol{\mathcal{Z}}$. Therefore, we get:

$$E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})'\big[\nabla_{\boldsymbol{\pi}}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\rho}(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) \otimes \boldsymbol{I}_{(k+2)}\big]\Big] = E\left[\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = -\Im_{\boldsymbol{\mathcal{Z}}} =$$

$$- E\Big[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big) + \nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})$$

$$+ \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})'\nabla_{\boldsymbol{\theta}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)' + \nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})'\nabla_{\boldsymbol{\pi}\boldsymbol{\pi}'}\ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{W}, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\big)\nabla_{\boldsymbol{\theta}}\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{Z})\Big]$$

and Eq. (25) becomes:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \xrightarrow{p} -\Im_{\boldsymbol{\mathcal{Z}}} \tag{26}$$

Using Lemmas A.10 and A.11, we can take the exact same steps as those used in the proof of Theorem 1(B) to show that equation (24) becomes:

$$\begin{aligned} &-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2 \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\pi}_N^*}(\overline{\boldsymbol{\theta}}, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ &= \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\partial \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}} + \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)\right] + o_p(\mathrm{N}^{-1/2}) \end{aligned}$$

and using (26) we have:

$$\begin{aligned} \sqrt{N}\big(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) &= \Im_{\boldsymbol{\mathcal{Z}}}^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}\left[\frac{\partial \ell_{\boldsymbol{\mathcal{Z}}}\big(\boldsymbol{w}_n, \boldsymbol{\theta}_0, \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}_n)\big)}{\partial\boldsymbol{\theta}} + \overline{D}_{\boldsymbol{\mathcal{Z}}}(\boldsymbol{z}_n)J_0\big(\boldsymbol{z}_n\big)^{-1}\Big(E\big[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}_n\big] - E\big[\boldsymbol{Y} \mid \boldsymbol{z}_n\big]\Big)\right] \\ &\quad + o_p(1) \\ &= \sqrt{N}\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p(1) \end{aligned}$$

where the last equality comes from the asymptotic linear representation of $\big(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)$ found in the proof of Theorem 1(B). This completes the proof. □

Proof of Corollary 2:

If the conditioning signals $\boldsymbol{\mathcal{Z}}$ are discrete and the conditions of the corollary are satisfied, proving Theorems 1 and 2 becomes significantly easier. In particular, we do not need to rely on Lemma A.1. The objects described in Lemmas 4.2 and 4.3 now converge uniformly in probability at speed $o_p(\mathrm{N}^{-1/2})$. Instead of relying on a result like Lemma A.1, these uniform convergence results can be proved employing standard dominance arguments (given the assumptions of the corollary). Employing the usual Taylor series expansions we can jump directly to a result equivalent to Lemma A.6 to show that:

$$\widehat{\boldsymbol{\pi}}_N^*(\boldsymbol{\theta}_0, \boldsymbol{z}) - \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) = \frac{J\big(\boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) \mid \boldsymbol{z}, \boldsymbol{\theta}_0\big)^{-1}}{\Pr(\boldsymbol{Z} = \boldsymbol{z})} \frac{1}{N} \sum_{n=1}^{N} \Big[E[\boldsymbol{Y} \mid \boldsymbol{x}_n, \boldsymbol{z}] - E[\boldsymbol{Y} \mid \boldsymbol{z}] \Big] \mathbb{1}\big\{\boldsymbol{z}_n = \boldsymbol{z}\big\} + o_p(\mathrm{N}^{-1/2})$$

for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Similarly, a result equivalent to Lemma A.9 allows to show that:

$$\nabla_{\boldsymbol{\theta}} \widehat{\boldsymbol{\pi}}_N^*(\boldsymbol{\theta}_0, \boldsymbol{z}) - \nabla_{\boldsymbol{\theta}} \boldsymbol{\pi}^*(\boldsymbol{\theta}_0, \boldsymbol{z}) = \frac{1}{\Pr(\boldsymbol{Z} = \boldsymbol{z})} \frac{1}{N} \sum_{n=1}^{N} \Big[\zeta(\boldsymbol{x}_n, \boldsymbol{z}) - E[\zeta(\boldsymbol{x}_n, \boldsymbol{z}) \mid \boldsymbol{z}] \Big] \mathbb{1}\big\{\boldsymbol{z}_n = \boldsymbol{z}\big\} + o_p(\mathrm{N}^{-1/2})$$

for some function $\underset{2\times(k+2)}{\zeta(\boldsymbol{x}, \boldsymbol{z})}$ and for all $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$. Using these results, the proofs of Theorems 1 and 2 proceed in a similar (but simpler) fashion as we did above. Namely, the proof relies once again on an application of the Central Limit Theorem for U-Statistics, without the need to employ Taylor series approximations for the expectations of the resulting U-statistics.

Proof of Corollary 3:

The proof would follow basically the exact same steps as the proof presented in the previous sections, starting with Lemma A.1 all the way through A.11. If the assumptions of the corollary are satisfied, Collomb and Hardle's result (the basis for Lemma A.1) are satisfied. The proof involves no important new considerations and can be safely omitted.

Before proceeding with the proof of Corollary 4 we begin by proving the following result, which is an extension of Lemma A.1.

Lemma A.12 *Let* $\{(\boldsymbol{X}_n, \boldsymbol{Z}_n)\}_{n=1}^N$ *be an iid sequence in* $\mathbb{R}^K \times \mathbb{R}^L$*, with* $\boldsymbol{X}_n$ *bounded with probability one. Suppose we have a kernel* $K : \mathbb{R}^L \to \mathbb{R}$ *that is symmetric, bounded and satisfies the conditions:* $\|u\| \cdot |K(u)| \to 0$ *as* $\|u\| \to \infty$*,* $\int K(u)du = 1$ *and the Lipschitz condition:* $\exists \gamma > 0$*,* $c_k < \infty$ *such that* $\big|K(u) - K(v)\big| \leq c_k \|u - v\|^\gamma$ $\forall u, v \in \mathbb{R}^L$*. Suppose the sequence* $\{h_N; N \in \mathbb{N}\}$ *is such that as* $N \to \infty$*:* $h_N \to 0$ *and* $N^{1-2\varepsilon} h_N^{2L} \to \infty$ *for some* $\varepsilon > 0$*. Let* $\eta : \mathbb{R}^K \times \mathbb{R}^L \times \mathbb{R}^P \to \mathbb{R}$ *be a continuously differentiable function that satisfies:*

$\left|\eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})\right| \le \overline{M} < \infty$, $\left\|\frac{\partial\eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})}{\partial\boldsymbol{t}}\right\| \le \overline{C}_1 < \infty$ *and* $\left\|\frac{\partial\eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})}{\partial\boldsymbol{z}}\right\| \le \overline{C}_2 < \infty$ *w.p.1 for all* $(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})$. *Now let:*

$$R_N(\boldsymbol{z},\boldsymbol{t}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N}\eta(\boldsymbol{X}_n,\boldsymbol{z},\boldsymbol{t})K\left(\frac{\boldsymbol{Z}_n-\boldsymbol{z}}{h_N}\right)$$

Take the set: $\boldsymbol{\mathcal{Z}}_{b_N} = \left\{\boldsymbol{z} \in \mathbb{R}^L : f_{\boldsymbol{Z}}(\boldsymbol{z}) \ge b_N\right\}$ *and define* $\boldsymbol{z}^*_{b_N} = \sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}} \left\|\boldsymbol{z}\right\|$. *We allow* $b_N \to 0$. *Suppose that* b_N *and* $f_{\boldsymbol{Z}}(\cdot)$ *are such that* $\log\left(\boldsymbol{z}^*_{b_N}\right) = o_p(N^{\varepsilon})$. *Now take any compact set* $\boldsymbol{G} \in \mathbb{R}^P$. *Then we have:*

$$\left(N^{1-\varepsilon}h_N^L\right)^{1/2} \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}\\ \boldsymbol{t}\in\boldsymbol{G}}} \left|R_N(\boldsymbol{z},\boldsymbol{t}) - ER_N(\boldsymbol{z},\boldsymbol{t})\right| = O_p(1) \quad \text{w.p.1}$$

Proof: We now have $\boldsymbol{\mathcal{Z}}_{b_N} = \left\{\boldsymbol{z} \in \mathbb{R}^L : f_{\boldsymbol{Z}}(\boldsymbol{z}) \ge b_N\right\}$. Note that $\boldsymbol{\mathcal{Z}}_{b_N}$ is compact for all $b_N > 0$ (otherwise $f_{\boldsymbol{Z}}(\cdot)$ would not be a well-behaved density). Now define $\boldsymbol{z}^*_{b_N} = \sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}} \left\|\boldsymbol{z}\right\|$ and $\boldsymbol{t}^* = \sup_{\boldsymbol{t}\in\boldsymbol{G}} \left\|\boldsymbol{t}\right\|$, both of which are finite by compactness of $\boldsymbol{\mathcal{Z}}_{b_N}$ and $\boldsymbol{G}$. Note that we have $\boldsymbol{\mathcal{Z}}_{b_N} \subseteq [-\boldsymbol{z}^*_{b_N}, \boldsymbol{z}^*_{b_N}]^L$ and $\boldsymbol{G} \subseteq [-\boldsymbol{t}^*, \boldsymbol{t}^*]^P$. Consider the following two collections of points in $\mathbb{R}$:

$$\left\{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{Q_N}\right\}, \quad \text{where } \boldsymbol{z}_k = -\boldsymbol{z}^*_{b_N} + \frac{k}{N} \text{ and } Q_N = 2\cdot\boldsymbol{z}^*_{b_N}N$$
$$\left\{\boldsymbol{t}_0, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_{S_N}\right\}, \quad \text{where } \boldsymbol{t}_\ell = -\boldsymbol{t}^* + \frac{\ell}{N} \text{ and } S_N = 2\cdot\boldsymbol{t}^*N$$

note that $\boldsymbol{z}_0 = -\boldsymbol{z}^*_{b_N}$, $\boldsymbol{z}_{Q_N} = \boldsymbol{z}^*_{b_N}$, $\boldsymbol{t}_0 = -\boldsymbol{t}^*$ and $\boldsymbol{t}_{S_N} = \boldsymbol{t}^*$. Define the following partitions in $\mathbb{R}^L$ and $\mathbb{R}^P$ respectively:

$$\mathcal{A}_N = \underbrace{\left\{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{Q_N}\right\} \times \left\{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{Q_N}\right\} \times \cdots \times \left\{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_{Q_N}\right\}}_{\text{L times}} \subset \mathbb{R}^L$$

$$\mathcal{G}_N = \underbrace{\left\{\boldsymbol{t}_0, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_{S_N}\right\} \times \left\{\boldsymbol{t}_0, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_{S_N}\right\} \times \cdots \times \left\{\boldsymbol{t}_0, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_{S_N}\right\}}_{\text{P times}} \subset \mathbb{R}^P$$

then, the partitions $\mathcal{A}_N$ and $\mathcal{G}_N$ satisfy:

$$\text{For all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}_{b_N}\text{: } \max_{\boldsymbol{v}\in\mathcal{A}_N} \left\|\boldsymbol{z}-\boldsymbol{v}\right\| \le \frac{\sqrt{L}}{N}. \quad \text{For all } \boldsymbol{t} \in \boldsymbol{G}\text{: } \max_{\boldsymbol{r}\in\mathcal{G}_N} \left\|\boldsymbol{t}-\boldsymbol{r}\right\| \le \frac{\sqrt{P}}{N}$$

The sets $\mathcal{A}_N$ and $\mathcal{G}_N$ have $M_N \equiv \left(2\boldsymbol{z}^*_{b_N} N\right)^L$ and $T_N \equiv \left(2\boldsymbol{t}^* N\right)^P$ elements respectively. Take any $(\boldsymbol{z},\boldsymbol{t}) \in \boldsymbol{\mathcal{Z}}_{b_N} \times \boldsymbol{G}$. From now on, we will denote:

$$\boldsymbol{z}_k = \underset{\boldsymbol{v}\in\mathcal{A}_N}{\operatorname{argmin}} \left\|\boldsymbol{z}-\boldsymbol{v}\right\| \quad \text{and} \quad \boldsymbol{t}_\ell = \underset{\boldsymbol{r}\in\mathcal{G}_N}{\operatorname{argmin}} \left\|\boldsymbol{t}-\boldsymbol{r}\right\|$$

Take any $\boldsymbol{t} \in G$ and $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}_{b_N}$. Let

$$R_N(\boldsymbol{z},\boldsymbol{t}) = \frac{1}{Nh_N^L}\sum_{n=1}^{N}\eta(\boldsymbol{X}_n,\boldsymbol{z},\boldsymbol{t})K_{h_N}(\boldsymbol{Z}_n-\boldsymbol{z})$$

The assumptions of the lemma imply that:

$$\left|R_N(\boldsymbol{z},\boldsymbol{t}) - R_N(\boldsymbol{z}_k,\boldsymbol{t}_\ell)\right| \le \frac{1}{h_N^L}\cdot\Big(\overline{K}\cdot\overline{C}_1\|\boldsymbol{t}-\boldsymbol{t}_\ell\| + \overline{M}c_k\|\boldsymbol{z}-\boldsymbol{z}_k\|^\gamma + \overline{K}\cdot\overline{C}_2\|\boldsymbol{z}-\boldsymbol{z}_k\|\Big) \quad \text{w.p.1}$$

Without loss of generality, assume $\gamma = 1$ in the Lipschitz condition for the kernel function[32]. Then, we get:

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}\\ \boldsymbol{t}\in\boldsymbol{G}}} \left|R_N(\boldsymbol{z},\boldsymbol{t}) - R_N(\boldsymbol{z}_k,\boldsymbol{t}_\ell)\right| \le \frac{\overline{C}}{Nh_N^L} \quad \text{w.p.1}$$

where $\overline{C} \equiv \overline{K}\cdot\overline{C}_1\sqrt{P} + \left(\overline{M}c_k + \overline{K}\cdot\overline{C}_2\right)\sqrt{L}$. Now let $U_N(\boldsymbol{z},\boldsymbol{t}) = R_N(\boldsymbol{z},\boldsymbol{t}) - ER_N(\boldsymbol{z},\boldsymbol{t})$. The result above implies that

$$\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}\\ \boldsymbol{t}\in\boldsymbol{G}}} \left|U_N(\boldsymbol{z},\boldsymbol{t})\right| \le \max_{\substack{k=1,\dots,M_N\\ \ell=1,\dots,T_N}} \left|U_N(\boldsymbol{z}_k,\boldsymbol{t}_\ell)\right| + \frac{2\overline{C}}{Nh_N^L}$$

By the assumptions of the lemma, there exist $\overline{D}_1 > 0$, $\overline{D}_2 > 0$ such that:

$$\operatorname{Var}\left[\eta(\boldsymbol{X}_n,\boldsymbol{z},\boldsymbol{t})K\left(\frac{\boldsymbol{Z}_n-\boldsymbol{z}}{h_N}\right)\right] \le \overline{D}_1 \ \forall\ \boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}, \boldsymbol{t}\in\boldsymbol{G}$$

$$\left|\eta(\boldsymbol{X},\boldsymbol{z},\boldsymbol{t})K\left(\frac{\boldsymbol{Z}-\boldsymbol{z}}{h_N}\right)\right| \le \overline{D}_2 \ \forall\ \boldsymbol{X}\in\mathbb{S}(\boldsymbol{X}), \boldsymbol{Z}\in\mathbb{S}(\boldsymbol{Z}), \boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}, \boldsymbol{t}\in\boldsymbol{G}$$

Now take any $\Delta > 0$. Using Bernstein's inequality we get:

$$\Pr\Big(\left|U_N(\boldsymbol{z},\boldsymbol{t})\right| > \Delta\Big) \le 2\cdot\exp\left(\frac{-Nh_N^L\Delta^2}{2\overline{D}_1 + 4\Delta\overline{D}_2/3}\right) \ \forall\ \boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}, \boldsymbol{t}\in\boldsymbol{G}$$

[32] For any $\gamma > 0$ we can always design the partition $\mathcal{A}_N$ so that $\min_{\boldsymbol{v}\in\mathcal{A}_N} \left\|\boldsymbol{z}-\boldsymbol{v}\right\|^\gamma \le \frac{\mathcal{C}}{N}$ for some finite constant $\mathcal{C} > 0$.

which yields:

$$\Pr\Big(\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\big|U_N(\boldsymbol{z},\boldsymbol{t})\big| > \Delta\Big) \leq 2\cdot\exp\left(\frac{-N^{\varepsilon}\Delta^2}{2\overline{D}_1 + 4\Delta\overline{D}_2/3}\right)\ \forall\ \boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}, \boldsymbol{t}\in\boldsymbol{G}$$

consequently:

$$\Pr\left(\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\max_{\substack{k=1,\ldots,M_N\\ \ell=1,\ldots,T_N}}\big|U_N(\boldsymbol{z}_k,\boldsymbol{t}_\ell)\big|\big| > \Delta\right) \leq \sum_{\ell=1}^{T_N}\sum_{k=1}^{M_N}\Pr\Big(\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\big|U_N(\boldsymbol{z}_k,\boldsymbol{t}_\ell)\big| > \Delta\Big)$$

$$\leq 2T_NM_N\cdot\exp\left(\frac{-N^{\varepsilon}\Delta^2}{2\overline{D}_1 + 4\Delta\overline{D}_2/3}\right)$$

Define $\widetilde{\Delta}\equiv\Delta - 2\overline{C}/\big(N^{1+\varepsilon}h_N^L\big)^{1/2}$. Then, these results show that:

$$\Pr\left(\big(N^{1-\varepsilon}h_N^L\big)^{1/2}\sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}\\ \boldsymbol{t}\in\boldsymbol{G}}}\big|U_N(\boldsymbol{z},\boldsymbol{t})\big| > \Delta\right) \leq 2T_NM_N\cdot\exp\left(\frac{-N^{\varepsilon}\widetilde{\Delta}^2}{2\overline{D}_1 + 4\widetilde{\Delta}\overline{D}_2/3}\right)$$

If $\log\left(\boldsymbol{z}^*_{b_N}\right) = o_p(\mathrm{N}^{\varepsilon})$ and $Nh_N\to\infty$ then $\log\big(2T_NM_N\big)/N^{\varepsilon}\to 0$ and $\widetilde{\Delta}\to\Delta$. Consequently $\left(\log\big(2T_NM_N\big)/N^{\varepsilon} - \widetilde{\Delta}^2\right)\to -\Delta^2$ and $T_NM_N\cdot\exp\left(\frac{-N^{\varepsilon}\widetilde{\Delta}^2}{2\overline{D}_1+4\widetilde{\Delta}\overline{D}_2/3}\right)\to 0$. Since $\Delta>0$ was arbitrary, this proves the result. □

Proof of Corollary 4:

The lemma assumes uniqueness of equilibrium everywhere in $\mathbb{S}(\boldsymbol{Z})$ (i.e, $\boldsymbol{\mathcal{Z}} = \mathbb{S}(\boldsymbol{Z})$). The proofs of Lemmas A.2-A.11 in $\boldsymbol{\mathcal{Z}}_{b_N}\times\boldsymbol{\Theta}$ follow from assumptions (S1)-(S4), Lemma A.12 and the assumption that $b_N^2\big(N^{1-2\varepsilon}h_N^{2L}\big)^{1/4}\to\infty$. From assumptions (S1)-(S4) we obtain that the biases of each of the semiparametric objects defined is still of order h_N^M uniformly in $\boldsymbol{\mathcal{Z}}_{b_N}\times\boldsymbol{\Theta}$. If assumption (S4.2) is satisfied then $b_N^2\big(N^{1-2\varepsilon}h_N^{2L}\big)^{1/4}\to\infty$ implies that $N^{1/4}h_N^M/b_N^2\to 0$. These facts are used to extend the results of Lemma A.2 as:

$$\sup_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}}\left|\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}) - f_{\boldsymbol{Z}}(\boldsymbol{z})\right| = o_p(\mathrm{N}^{-1/4}),\qquad \sup_{\substack{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}_{b_N}\\ \boldsymbol{\theta}_p\in\boldsymbol{B}\\ \pi_{-p}\in\boldsymbol{A}}}\left|\widehat{\varphi}_{p_N}(\pi_{-p}\mid\boldsymbol{z},\boldsymbol{\theta}_p) - \varphi_p(\pi_{-p}\mid\boldsymbol{z},\boldsymbol{\theta}_p)\right| = o_p(\mathrm{N}^{-1/4})$$

and so on, for the rest of the objects defined in Lemma A.2. The rest of the lemmas - including Lemmas 4.2-4.3- follow from here. To complete the proof of the corollary, we have to show that the results of Theorems 1 and 2. Given that the results of Lemmas A.2-A.11 can be extended to the set $\boldsymbol{\mathcal{Z}}_{b_N} \times \boldsymbol{\Theta}$, all that is left is to characterize the asymptotic behavior of the trimming function $\mathbb{1}\{f_{\boldsymbol{Z}_N}(\boldsymbol{z}) > b_N\}$. To do this, note first that $b_N^2\left(N^{1-2\varepsilon}h_N^{2L}\right)^{1/4} \to \infty$ implies that $Nh_N^L b_N^2/\log N \to \infty$ and $b_N/h_N^M \to \infty$. These results along with assumption (S2.1-S2.2) and (S4) imply that all the conditions of Lemma 25 in Ichimura (2004) are satisfied, and we get:

$$\Pr\left(\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n) > b_N\} - \mathbb{1}\{f_{\boldsymbol{Z}}(\boldsymbol{z}_n) > b_N\} \neq 0 \ \text{ for at least one } \boldsymbol{z}_n\right) \to 0$$

Therefore, the asymptotic properties of $N^{-1}\sum_{n=1}^{N} \log \mathcal{F}\left(\boldsymbol{y}_n \mid \boldsymbol{x}_n, \boldsymbol{z}_n, \boldsymbol{\theta}\right)\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n) > b_N\}$ are the same as those of $N^{-1}\sum_{n=1}^{N} \log \mathcal{F}\left(\boldsymbol{y}_n \mid \boldsymbol{x}_n, \boldsymbol{z}_n, \boldsymbol{\theta}\right)\mathbb{1}\{f_{\boldsymbol{Z}}(\boldsymbol{z}_n) > b_N\}$. Since we have: $\{\boldsymbol{z} \in \mathbb{R}^L : f_{\boldsymbol{Z}}(\boldsymbol{z}) > b_N\} \subseteq \boldsymbol{\mathcal{Z}}_{b_N}$, then the fact that Lemmas A.2-A.11 hold in $\boldsymbol{\mathcal{Z}}_{b_N} \times \boldsymbol{\Theta}$, it follows that the results of Theorems 1 and 2 hold when the trimming function is $\mathbb{1}\{\widehat{f}_{\boldsymbol{Z}_N}(\boldsymbol{z}_n) > b_N\}$. Also, $b_N \to 0$ implies that $\boldsymbol{\mathcal{Z}}_{b_N} \to \mathbb{S}(\boldsymbol{Z})$ and the asymptotic distributions of $\widehat{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\theta}}$ do not depend on any trimming set, as was claimed. □

References

[1] Ai, C. and X Chen (2003) "Efficient estimation of models with conditional moment restrictions containing unknown functions" Econometrica 71: pp. 1795-1843.

[2] Collomb, G. and W. Hardle (1986) "Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations", Stochastic Processes and their Applications 23: pp. 77-89.

[3] Ichimura, H. "Computation of asymptotic distribution for semiparametric GMM estimators" Unpublished manuscript.

[4] Mas-Collel, A., Whinston, M. and J.R Green (1995) Microeconomic Theory (Oxford: Oxford University Press).

[5] Newey, W.K (1990) "Semiparametric efficiency bounds" Journal of Applied Econometrics 5: pp. 99-135

[6] Newey, W.K (1994) "The asymptotic variance of semiparametric estimators" Econometrica 62: pp. 1349-1382.

[7] Newey, W.K and D. McFadden (1994) "Large sample estimators and hypothesis testing" in: R.Engle and D.McFadden, eds., Handbook of Econometrics, Vol. 4 (North Holland, Amsterdam) pp. 2113-2245.

[8] Pagan, A. and A. Ullah (1999) Nonparametric Econometrics (Cambridge University Press, Cambridge).

[9] Powell, J.L, J.H Stock and T.M Stoker (1989) "Semiparametric estimation of index coefficients" Econometrica 57: pp. 1403-1430.