Supplemental Appendix for

"Estimating Demand for Differentiated Products with Zeroes in Market Share Data"

Amit Gandhi, Zhentong Lu, Xiaoxia Shi*

In this supplemental appendix, we present supporting materials for "Estimating Demand for Differentiated Products with Zeroes in Market Share Data" (hereafter "main text"). The supplemental appendix is organized as follows:

Section A provides further illustration of the power law pattern in homicide and international trade data sets. This section complements the illustration in Section 2 in the main text.

Section B gives the proofs of Lemmas 1 and 2 presented in Sections 4 and 5 in the main text, respectively. Lemma 1 establishes the validity of the bounds for the general model. Lemma 2 proves that the bounds collapse for the dominant products.

Section C proves Theorems 1 and 2 presented in Sections 6 and 7 in the main text, respectively. Theorem 1 establishes the consistency of our proposed estimator and Theorem 2 proves the asymptotic normality.

Section D proves a lemma that establishes Assumption 1 in Section 4 of the main text for the random coefficient logit model.

Section E provides analytical evidence for the bound validity in Section 3 in the main text. The two lemmas presented here reinforce the numerical proof given in Section 3.

A Further Illustrations of Zipf's Law

In Figure 3 we illustrate this regularity using data from the two other applications that were mentioned in Section 2: homicide rates and international trade flows. The left hand graph shows the annual murder rate (per 10,000 people) for each county in the US from 1977-1992 (for details about the data see ?). The right hand side graph shows the import trade flows (measured in millions of US dollars) among 160 countries that have a regional trade agreement in the year 2006 (for details about the data see ?). In each of these two

* Corresponding author: xshi@ssc.wisc.edu

cases we see the characteristic pattern of Zipf's law - a sharp decay in the frequency for large outcomes and a large mass near zero (with a mode at zero in each case).

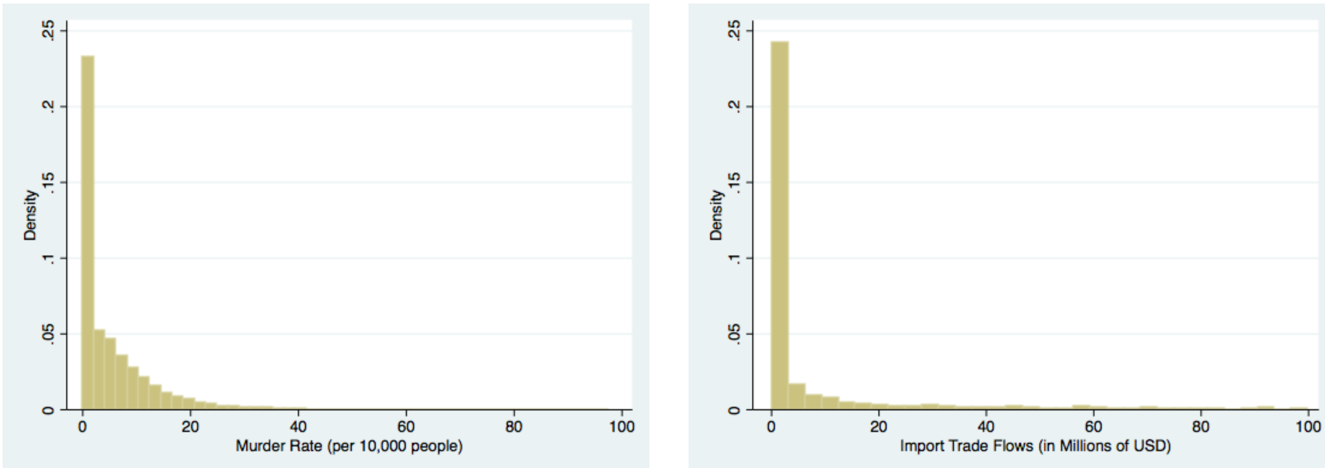

Figure 3: Zipf's Law in Crime and Trade Data

B Proof of Lemmas 1 and 2

Proof of Lemma 1. We start with the case where Assumption 1 holds. We show the argument for the upper bound only because the lower bound is analogous. Consider the derivation

$$\hat{\delta}^u_{jt}(s_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0) = [\log((n_t s_{jt} + \iota_u)/n_t) - \log(\pi_{jt})] + [\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0)]. \quad (42)$$

Let $e^u_{jt} = \check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0)$. Then by Assumption 1(b),

$$E[\hat{\delta}^u_{jt}(s_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0) - e^u_{jt}|\pi_{jt}, z_{jt}] = E[\log((n_t s_{jt} + \iota_u)/n_t) - \log(\pi_{jt})|\pi_{jt}, z_{jt}] \geq 0. \quad (43)$$

Since $E[\xi_{jt}|z_{jt}] = 0$, we have $E[\delta_{jt}(\pi_t, \lambda_0) - x_{jt}\beta_0|z_{jt}] = 0$. Thus,

$$E[\hat{\delta}^u_{jt}(s_t, \lambda_0) - x_{jt}\beta_0 - e^u_{jt}|z_{jt}] \geq 0. \quad (44)$$

Let u_t stand for $\frac{n_t^{1/2}}{T^{1/4}J_t^{1/2}}$. Now we show that $\sup_{j,t} u_t|e^u_{jt}| = O_p(1)$. Consider the derivation:

$$\sup_{j,t} u_t|e^u_{jt}| = \sup_{j,t} u_t|\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0)|$$

$$
\begin{aligned}
&\leq \sup_{j,t} \frac{|\check{\delta}_{jt}(\tilde{s}_t,\lambda_0) - \check{\delta}_{jt}(\pi_t,\lambda_0)|}{\|\tilde{s}_t - \pi_t\|_f \sqrt{J_t}} \sup_t \sqrt{J_t} u_t \|\tilde{s}_t - \pi_t\|_f \\
&\leq O_p(1) \left(\sup_t u_t \sqrt{J_t} \|\tilde{s}_t - s_t\|_f + \sup_{t=1} u_t \sqrt{J_t} \|s_t - \pi_t\|_f \right) \\
&= o_p(1) + O_p(1) \sup_t u_t \sqrt{J_t} \|s_t - \pi_t\|_f, \qquad (45)
\end{aligned}
$$

where the second inequality holds by Assumption 1(a) and the second equality holds by Assumption 1(b). To bound $u_t\sqrt{J_t}\|s_t - \pi_t\|_f$, note that $n_t s_{jt}$ is a binomial random variable with parameters (n_t, π_{jt}). Thus,

$$
\begin{aligned}
\Pr\left(\sup_{t=1,\dots,T} u_t J_t^{1/2} \|s_t - \pi_t\|_f > \varepsilon \right) &\leq \sum_{t=1}^{T} \Pr(u_t J_t^{1/2} \|s_t - \pi_t\|_f > \varepsilon) \\
&\leq \sum_{t=1}^{T} \frac{128 u_t^4 J_t^2 (3n_t^2 + n_t)}{n_t^4 \varepsilon^4} \\
&\leq \frac{512}{\varepsilon^4}, \qquad (46)
\end{aligned}
$$

where the second inequality holds by Lemma 8. The expression in the last line does not depend on T and it can be made arbitrarily small by making ε big. This shows that

$$
\sup_{t=1,\dots,T} u_t J_t^{1/2} \|s_t - \pi_t\|_f = O_p(1) \qquad (47)
$$

which implies $\sup_{j,t} u_t |e_{jt}^u| = O_p(1)$ when combined with (45).

Now we move on to the case where Assumption 2 holds instead. In this case, the upper bound and the lower bound need slightly different arguments. For the upper bound, consider the derivation:

$$
\begin{aligned}
\hat{\delta}_{jt}^u(s_t,\lambda_0) - \delta_{jt}(\pi_t,\lambda_0) &= \delta_{jt}(\tilde{s}_{jt},\lambda_0) - \delta_{jt}(\pi_t,\lambda_0) + \log((n_t s_{jt} + \iota_u)/n_t) - \log(\tilde{s}_{jt}) \\
&\geq \delta_{jt}(\tilde{s}_{jt},\lambda_0) - \delta_{jt}(\pi_t,\lambda_0), \qquad (48)
\end{aligned}
$$

where the inequality holds because $\tilde{s}_{jt} = s_{jt} + 1/n_t$ and $\iota_u > 1$ both by Assumption 2(c). Equation (48) combined with Assumption 2(b) implies the first line of (25) with $e_{jt}^u = 0$.

For the lower bound, consider the derivation:

$$
\hat{\delta}_{jt}^\ell(s_t,\lambda_0) - \delta_{jt}(\pi_t,\lambda_0) = \check{\delta}_{jt}(\tilde{s}_t,\lambda_0) - \check{\delta}_{jt}(\pi_t,\lambda_0) + \log((n_t s_{jt} + \iota_\ell)/n_t) - \log(\pi_{jt}) \qquad (49)
$$

By $\iota_\ell \le \bar{\iota}_\ell$ (Assumption 2(c)) and the definition of $\bar{\iota}_\ell$, we have

$$E[\log((n_t s_{jt} + \iota_\ell)/n_t] - \log(\pi_{jt})|n_t, \pi_t] \le 0. \tag{50}$$

This combined with Assumption 2(a) implies the second line of (25) with $e^\ell_{jt} = 0$. □

Proof of Lemma 2. Observe that

$$\begin{aligned} n_t|\hat{\delta}^u_{jt}(s_t,\lambda) - \hat{\delta}^\ell_{jt}(s_t,\lambda)| &= n_t|\log(s_{jt} + \iota_u/n_t) - \log(s_{jt} + \iota_\ell/n_t)| \\ &\le \frac{1}{s_{jt} + \iota_\ell/n_t}(\iota_u - \iota_\ell), \end{aligned} \tag{51}$$

using the concavity of the logarithm function. Thus

$$\begin{aligned} &\sup_{j=1,\dots,J_t;t=1,\dots,T} \sup_\lambda n_t|\hat{\delta}^u_{jt}(s_t,\lambda) - \hat{\delta}^\ell_{jt}(s_t,\lambda)|1\{z_{jt} \in \mathcal{Z}_0\} \\ &\le \sup_{j=1,\dots,J_t;t=1,\dots,T} \frac{\iota_u - \iota_\ell}{s_{jt} + \iota_\ell/n_t} 1\{z_{jt} \in \mathcal{Z}_0\} \\ &\le \frac{\iota_u - \iota_\ell}{\inf_{j=1,\dots,J_t;t=1,\dots,T}\{(s_{jt} + \iota_\ell/n_t)1\{z_{jt} \in \mathcal{Z}_0\} + 1\{z_{jt} \notin \mathcal{Z}_0\}\}}. \end{aligned} \tag{52}$$

The denominator of (52) is greater than or equal to

$$\inf_{j=1,\dots,J_t;t=1,\dots,T}\{\pi_{jt}1\{z_{jt} \in \mathcal{Z}_0\} + 1\{z_{jt} \notin \mathcal{Z}_0\}\} - \sup_{j=1,\dots,J_t;t=1,\dots,T} |\pi_{jt} - s_{jt} - \iota_\ell/n_t|1\{z_{jt} \in \mathcal{Z}_0\}. \tag{53}$$

Consider that

$$\begin{aligned} &\Pr\left(\inf_{j=1,\dots,J_t;t=1,\dots,T}\{\pi_{jt}1\{z_{jt} \in \mathcal{Z}_0\} + 1\{z_{jt} \notin \mathcal{Z}_0\}\} < \underline{\varepsilon}_0\right) \\ &\le \sum_{j=1,\dots,J_t;t=1,\dots,T} \Pr(\pi_{jt}1\{z_{jt} \in \mathcal{Z}_0\} + 1\{z_{jt} \notin \mathcal{Z}_0\} < \underline{\varepsilon}_0) \\ &= \sum_{j=1,\dots,J_t;t=1,\dots,T} \Pr(\pi_{jt} < \underline{\varepsilon}_0 | z_{jt} \in \mathcal{Z}_0) P(z_{jt} \in \mathcal{Z}_0) \\ &= 0, \end{aligned} \tag{54}$$

where the first equality holds since $1 \ge \underline{\varepsilon}_0$, and the second equality holds by Assumption 3.

Also consider the derivation:

$$\Pr\left(\sup_{j=1,\dots,J_t;t=1,\dots,T} |\pi_{jt} - s_{jt} - \iota_\ell/n_t|1\{z_{jt} \in \mathcal{Z}_0\} > \underline{\varepsilon}_0/2\right)$$

$$
\begin{aligned}
&\le \Pr\left(\sup_{t=1,\dots,T}\|s_t-\pi_t\|_f > \underline{\varepsilon}_0/4\right)+\Pr\left(\sup_{t=1,\dots,T}\iota_\ell/n_t > \underline{\varepsilon}_0/4\right)\\
&\le \sum_{t=1}^{T}\Pr(\|s_t-\pi_t\|_f > \underline{\varepsilon}_0/4)+o(1)\\
&\le \sum_{t=1}^{T}\frac{128(3n_t^2+n_t)\times 4^4}{n_t^4\underline{\varepsilon}_0^4}+o(1)\\
&\to 0, \qquad (55)
\end{aligned}
$$

where the first inequality holds by the triangular inequality, the second inequality and the convergence hold Assumption 4(g), and the third inequality holds by Lemma 8.

Equations (52), (54), and (55) together imply that

$$
\Pr\left(\sup_{j=1,\dots,J_t;t=1,\dots,T}\sup_{\lambda} n_t|\hat{\delta}^u_{jt}(s_t,\lambda)-\hat{\delta}^\ell_{jt}(s_t,\lambda)|1\{z_{jt}\in\mathcal{Z}_0\} > \frac{\iota_u-\iota_\ell}{\underline{\varepsilon}_0/2}\right)\to 0. \qquad (56)
$$

This proves the lemma. □

C Proofs of the Theorems

In this section, we prove the theorems that establish the consistency and the asymptotic normality of our proposed estimator.

C.1 Proof of Theorem 1: Consistency

The proof of Theorem 1 uses the following lemma which is proved in Section C.2 below.

Lemma 3. *Suppose that either Assumption 1 holds and $T^{-1}\sum_{t=1}^{T}J_t^2/\bar{J}_T^2$ is bounded, or Assumption 2 holds and $\sup_{t=1,\dots,T}J_t$ is bounded. Also suppose that Assumptions 3-6 hold. Then,*

(i) $\sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}_0}|\bar{m}^u_T(\theta,g)-\bar{m}_T(\theta,g)|=o_p(1)$ *and* $\sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}_0}|\bar{m}^\ell_T(\theta,g)-\bar{m}_T(\theta,g)|=o_p(1)$.

(ii) $\sum_{g\in\mathcal{G}}\mu(g)[\bar{m}^u_T(\theta_0,g)]_-^2=o_p(1)$ *and* $\sum_{g\in\mathcal{G}}\mu(g)[\bar{m}^\ell_T(\theta_0,g)]_+^2=o_p(1)$.

Proof of Theorem 1. First note that $\widehat{Q}_T(\theta_0)=\sum_{g\in\mathcal{G}}\mu(g)[\bar{m}^u_T(\theta_0,g)]_-^2+\sum_{g\in\mathcal{G}}\mu(g)[\bar{m}^\ell_T(\theta_0,g)]_+^2$. Thus, Lemma 3(b) implies that

$$
\widehat{Q}_T(\theta_0)=o_p(1). \qquad (57)
$$

Define an auxiliary criterion function:

$$\widehat{Q}_{0,T}(\theta) = \sum_{g \in \mathcal{G}_0} \left\{ \left([\bar{m}_T^u(\theta, g)]_-^2 + \left[\bar{m}_T^\ell(\theta, g)\right]_+^2 \right) \mu(g) \right\}.$$

Below we show that

$$\sup_{\theta \in \Theta} \left| \sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)} \right| = o_p(1). \tag{58}$$

Consider an arbitrary $c > 0$. The theorem is implied by the following derivation:

$$\begin{aligned}
\Pr\left(\|\widehat{\theta}_T^s - \theta_0^s\| > c\right) &\leq \Pr\left(\sqrt{\widehat{Q}_T^*(\widehat{\theta}_T)} \geq \sqrt{C(c)}\right) \\
&= \Pr\left(\sqrt{\widehat{Q}_T^*(\widehat{\theta}_T)} - \sqrt{\widehat{Q}_{0,T}(\widehat{\theta}_T)} + \sqrt{\widehat{Q}_{0,T}(\widehat{\theta}_T)} \geq \sqrt{C(c)}\right) \\
&\leq \Pr\left(\sup_{\theta \in \Theta} |\sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)}| + \sqrt{\widehat{Q}_{0,T}(\widehat{\theta}_T)} \geq \sqrt{C(c)}\right) \\
&\leq \Pr\left(\sup_{\theta \in \Theta} |\sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)}| + \sqrt{\widehat{Q}_T(\widehat{\theta}_T)} \geq \sqrt{C(c)}\right) \\
&\leq \Pr\left(\sup_{\theta \in \Theta} |\sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)}| + \sqrt{\widehat{Q}_T(\theta_0)} \geq \sqrt{C(c)}\right) \\
&\leq \Pr\left(\sup_{\theta \in \Theta} |\sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)}| \geq \sqrt{C(c)}/2\right) + \Pr\left(\widehat{Q}_T(\theta_0) \geq C(c)/4\right) \\
&\to 0,
\end{aligned} \tag{59}$$

where the first inequality holds by Assumption 7, the third inequality holds because $\widehat{Q}_T(\widehat{\theta}_T)$ differs from $\widehat{Q}_{0,T}(\widehat{\theta}_T)$ only in that the former takes the summation over a larger range, the fourth inequality holds because $\widehat{Q}_T(\widehat{\theta}_T) \leq \widehat{Q}_T(\theta_0)$ by the definition of $\widehat{\theta}_T$ and the convergence holds by (57) and (58).

Now we show (58). Consider the derivation

$$\begin{aligned}
&\sup_{\theta \in \Theta} \left| \sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)} \right| \\
=&\sup_{\theta \in \Theta} \left| \sqrt{\sum_{g \in \mathcal{G}_0} \mu(g) \left\{ [\bar{m}_T^u(\theta, g)]_-^2 + [\bar{m}_T^\ell(\theta, g)]_+^2 \right\}} - \sqrt{\sum_{g \in \mathcal{G}_0} \mu(g) \left\{ \bar{m}_T(\theta, g)^2 \right\}} \right| \\
\leq&\sup_{\theta \in \Theta} \left| \sqrt{\sum_{g \in \mathcal{G}_0} \mu(g) \left\{ \left(\sqrt{[\bar{m}_T^u(\theta, g)]_-^2 + [\bar{m}_T^\ell(\theta, g)]_+^2} - |\bar{m}_T(\theta, g)| \right)^2 \right\}} \right|
\end{aligned}$$

$$\leq \sup_{\theta\in\Theta}\left|\sqrt{\sum_{g\in\mathcal{G}_0}\mu(g)\left\{([\bar{m}_T^u(\theta,g)]_- - [\bar{m}_T(\theta,g)]_-)^2 + ([\bar{m}_T^\ell(\theta,g)]_+ - [\bar{m}_T(\theta,g)]_+)^2\right\}}\right|$$

$$\leq \sup_{\theta\in\Theta}\left|\sqrt{\sum_{g\in\mathcal{G}_0}\mu(g)\left\{(\bar{m}_T^u(\theta,g) - \bar{m}_T(\theta,g))^2 + (\bar{m}_T^\ell(\theta,g) - \bar{m}_T(\theta,g))^2\right\}}\right|$$

$$\leq \sqrt{\sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}_0}|\bar{m}_T^u(\theta,g) - \bar{m}_T(\theta,g)|^2 + \sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}_0}|\bar{m}_T^\ell(\theta,g) - \bar{m}_T(\theta,g)|^2}$$

$$\to_p 0, \tag{60}$$

where the first inequality holds by the triangular inequality for the norm

$$\|a(\cdot)\| := \sqrt{\sum_{g\in\mathcal{G}_0}\mu(g)a(g)^2 / \sum_{g\in\mathcal{G}_0}\mu(g)},$$

the second inequality holds by the triangular inequality for the Euclidean norm, the third inequality holds because $|[x]_- - [y]_-| \leq |x-y|$ and $[x]_+ = [-x]_-$, and the fourth inequality holds because $\mu : \mathcal{G} \to [0,1]$ is a probability measure on $\mathcal{G}$ and $\mathcal{G}_0 \subseteq \mathcal{G}$, and the convergence holds by Lemma 3(a). Therefore (58) is proved. □

C.2 Proof of Lemma 3

Proof of Lemma 3. First we show part (a). Let $\sup_{j,t:z_{jt}\in\mathcal{Z}_0}$ abbreviate $\sup_{t=1,\ldots,T}\sup_{j=1,\ldots,J_t:z_{jt}\in\mathcal{Z}_0}$. Consider the derivation:

$$\begin{aligned}
&\sup_{\theta\in\Theta}\sup_{g\in\mathcal{G}_0}|\bar{m}_T^u(\theta,g) - \bar{m}_T(\theta,g)| \\
&= \sup_{\lambda\in\Lambda}\sup_{g\in\mathcal{G}_0}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\hat{\delta}_{jt}^u(s_t,\lambda) - \delta_{jt}(\pi_t,\lambda))g(z_{jt})\right| \\
&\leq \sup_{\lambda\in\Lambda}\sup_{j,t:z_{jt}\in\mathcal{Z}_0}|\hat{\delta}_{jt}^u(s_t,\lambda) - \delta_{jt}(\pi_t,\lambda)| \\
&\leq \sup_{j,t:z_{jt}\in\mathcal{Z}_0}|\log(s_{jt}+\iota_u/n_t) - \log(\tilde{s}_{jt})| + \sup_{\lambda\in\Lambda}\sup_{j,t:z_{jt}\in\mathcal{Z}_0}|\delta_{jt}(\tilde{s}_t,\lambda) - \delta_{jt}(\pi_t,\lambda)|,
\end{aligned}$$

where the first inequality holds by the definition of $\mathcal{G}_0$.

Assumptions 4(f) and $0 < \iota_u < \infty$ (Assumption 1(b) or Assumption 2(b)) together imply that $\sup_{j,t:z_{jt}\in\mathcal{Z}_0}|s_{jt}+\iota_u/n_t - \tilde{s}_{jt}| \to_p 0$. Also, by equation (47) and Assumption 4(f)

$$\frac{n_t^{1/2}}{T^{1/4}}\sup_t\|\tilde{s}_t - \pi_t\|_f = O_p(1). \tag{61}$$

These, and Assumptions 3, 4(g), and 5(a) together imply that

$$\Pr\left(\inf_{j,t:z_{jt}\in\mathcal{Z}_0}\pi_{jt}\wedge\pi_{0t}>\underline{\varepsilon}_0,\inf_{j,t:z_{jt}\in\mathcal{Z}_0}s_{jt}+\iota_u/n_t>\underline{\varepsilon}_0/2,\inf_{j,t:z_{jt}\in\mathcal{Z}_0}\tilde{s}_{jt}\wedge\tilde{s}_{0t}>\underline{\varepsilon}_0/2\right)\to 1.$$

This combined with (61), Assumptions 4(g), and Assumption 6 implies that

$$\sup_{\lambda\in\Lambda}\sup_{j,t:z_{jt}\in\mathcal{Z}_0}|\delta_{jt}(\tilde{s}_t,\lambda)-\delta_{jt}(\pi_t,\lambda)|\to_p 0.$$

Also, we have

$$\sup_{j,t:z_{jt}\in\mathcal{Z}_0}(|\log(s_{jt}+\iota_u/n_t)-\log(\tilde{s}_{jt})|\to_p 0.$$

because the logarithm function is uniformly continuous on the closed interval $[\underline{\varepsilon}_0/2,1]$. Therefore, the first convergence in Lemma 3(a) holds. The second convergence holds by analogous arguments.

Now we show part (b). We separate the two cases, one where Assumption 1 is satisfied and the other where Assumption 2 is satisfied and J_t is bounded.

Case 1: Assumption 1 is satisfied. In this case, the arguments for the first convergence and the second convergence in part (b) are exactly analogous. Thus, we only discuss the first. Consider the derivation:

$$\begin{aligned}\bar{m}_T^u(\theta_0,g)&=\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\hat{\delta}_{jt}^u(s_t,\lambda_0)-x_{jt}'\beta_0)g(z_{jt})\\&\geq\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\xi_{jt}g(z_{jt})+\\&\quad\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt}+\underline{\iota}_u/n_t)-\log(\pi_{jt}))g(z_{jt})+\\&\quad\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t,\lambda_0)-\check{\delta}_{jt}(\pi_t,\lambda_0))g(z_{jt}),\end{aligned}\tag{62}$$

where the inequality holds because $\iota_u\geq\underline{\iota}_u$, $\xi_{jt}=\delta_{jt}(\pi_t,\lambda_0)-x_{jt}'\beta_0$ and $\check{\delta}(\cdot,\lambda)=\delta(\cdot,\lambda)-\log(\cdot_j)$. We analyze the three summands one by one. For the first summand, observe that $E[\xi_{jt}^2]\leq M$ by Assumption 4(e). We can then apply Lemma 7 in Appendix C.4 (with

$w_{jt} = \xi_{jt}$) and get,

$$E \sup_{g \in \mathcal{G}} \left| \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \{\xi_{jt} g(z_{jt}) - E[\xi_{jt} g(z_{jt})]\} \right|^2 \leq \frac{CM \sum_{t=1}^{T} J_t^2}{T^2 \bar{J}_T^2} = O(T^{-1}), \tag{63}$$

where the equality holds because we assume $T^{-1} J_t^2 / \bar{J}_T^2$ is bounded when Assumption 1 holds. Lemma 7 applies due to Assumptions 4(c)-(e). Also, by Assumption 4(c), $E[\xi_{jt} g(z_{jt})] = 0$. This and (63) together imply that

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \xi_{jt} g(z_{jt}) \right| = O_p(T^{-1/2}). \tag{64}$$

Similar arguments apply to the second summand in (62) and yields

$$\begin{aligned} & E \sup_{g \in \mathcal{G}} \left| \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt})) g(z_{jt}) - E[(\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt})) g(z_{jt})] \right|^2 \\ \leq & \frac{C \sum_{t=1}^{T} J_t^2}{T^2 \bar{J}_T^2} \max_{j,t} E[(\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt}))^2] \\ \leq & \frac{2C \sum_{t=1}^{T} J_t^2}{T^2 \bar{J}_T^2} \max_{t} [|\log(\underline{\iota}_u / n_t)|^2 + |\log(\underline{\varepsilon}_1 / n_t)|^2] \\ \leq & \frac{4C \sum_{t=1}^{T} J_t^2 (2(\log \overline{n}_T)^2 + (\log \underline{\iota}_u)^2 + (\log \underline{\varepsilon}_1)^2)}{T^2 \bar{J}_T^2} \qquad (65) \\ \to & 0, \end{aligned}$$

where the second inequality holds by $s_{jt} \in [0, 1]$ and Assumption 5(b) and the convergence holds by Assumptions 4(f) and the boundedness of $T^{-1} \sum_{t=1}^{T} J_t^2 / \bar{J}_t^2$. By the definition of $\underline{\iota}_u$, we have $E[(\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt})) | \pi_{jt}, z_{jt}] \geq 0$, which then implies that $E[(\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt})) g(z_{jt})] \geq 0$ for all $g \in \mathcal{G}$. Therefore, for any $c > 0$,

$$\lim_{T \to \infty} \Pr \left(\inf_{g \in \mathcal{G}} \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\log(s_{jt} + \underline{\iota}_u / n_t) - \log(\pi_{jt})) g(z_{jt}) < -c \right) = 0. \tag{66}$$

For the third summand in (62), consider the derivation

$$\begin{aligned} \sup_{g \in \mathcal{G}} \left| \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0)) g(z_{jt}) \right| & \leq \sup_{j,t} |\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0)| \\ & \to_p 0, \end{aligned} \tag{67}$$

by (61) and Assumptions 1(a) and 4(g). Finally, (62), (64), (66), and (67) combined imply that for any $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \bar{m}^u_T(\theta_0, g) < -c\right) = 0,$$

which then implies the first convergence in Lemma 3(b) since $[\bar{m}^u_T(\theta_0, g)]_- = \max\{0, -\bar{m}^u_T(\theta_0, g)\}$.

Case 2: Assumption 2 is satisfied. We begin with the first convergence in Lemma 3(b). Consider the decomposition:

$$\begin{aligned}
\bar{m}^u_T(\theta_0, g) =& \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\hat{\delta}^u_{jt}(s_t, \lambda_0) - x'_{jt}\beta_0)g(z_{jt}) \\
=& \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\xi_{jt}g(z_{jt})+ \\
& \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt} + \iota_u/n_t) - \log(\tilde{s}_{jt}))g(z_{jt})+ \\
& \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))g(z_{jt}). \qquad (68)
\end{aligned}$$

The first summand is $O_p(T^{-1/2})$ uniformly over $g \in \mathcal{G}$ by (64). The second summand is nonnegative almost surely because $\tilde{s}_{jt} = s_{jt} + 1/n_t$ and $\iota_u \geq 1$ (Assumption 2(d)). For the third summand, similar to (65), we get for some generic constant C,

$$\begin{aligned}
& E\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))g(z_{jt}) - E[(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))g(z_{jt})]\right|^2 \\
\leq& \frac{C\sum_{t=1}^{T} J_t^2}{T^2\bar{J}_T^2}\max_{j,t} E[(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))^2] \\
\leq& \frac{2CC_0\sum_{t=1}^{T} J_t^2\log(\overline{n}_T)^2}{T^2\bar{J}_T^2} \\
\to& 0, \qquad (69)
\end{aligned}$$

where the second inequality holds by Assumption 2(d) also using Assumptions 2(c) and 5(b), and the convergence holds by Assumption 4(g) and the boundedness of $\sup_{t=1,\ldots,T} J_t$. Moreover, Assumption 2(b) implies that $E[(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))g(z_{jt})] \geq 0$. This combined with (69) implies that, for any $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \frac{1}{T\bar{J}}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t, \lambda_0) - \delta_{jt}(\pi_t, \lambda_0))g(z_{jt}) < -c\right) = 0. \qquad (70)$$

This combined with the arguments for the first two summands of (62) above yields: for any $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \bar{m}^u_T(\theta_0, g) < -c\right) = 0,$$

which then implies the first convergence in Lemma 3(b) because $[\bar{m}^u_T(\theta_0, g)]_- = \max\{0, -\bar{m}^u_T(\theta_0, g)\}$.

Now we show the second convergence in Lemma 3(b) for Case 2. Note that

$$\begin{aligned}
\bar{m}^\ell_T(\theta_0, g) &= \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\hat{\delta}^\ell_{jt}(s_t, \lambda_0) - x'_{jt}\beta_0)g(z_{jt}) \\
&\leq \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\xi_{jt}g(z_{jt}) + \\
&\quad \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt} + \bar{\iota}_\ell/n_t) - \log(\pi_{jt}))g(z_{jt}) + \\
&\quad \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0))g(z_{jt}),
\end{aligned} \tag{71}$$

where the inequality holds because $\iota_\ell \leq \bar{\iota}_\ell$ by Assumption 2(d). The first summand is $O_p(T^{-1/2})$ by (64). Arguments analogous to those for (66) apply to the second summand to yield, for any $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt} + \bar{\iota}_\ell/n_t) - \log(\pi_{jt}))g(z_{jt}) > c\right) = 0. \tag{72}$$

For the third summand in (71), we can apply the same arguments as those for (70) where we use Assumption 2(a) in place of Assumption 2(b). Such arguments yield, for all $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t, \lambda_0) - \check{\delta}_{jt}(\pi_t, \lambda_0))g(z_{jt}) > c\right) = 0. \tag{73}$$

Therefore, for any $c > 0$,

$$\lim_{T\to\infty} \Pr\left(\inf_{g\in\mathcal{G}} \bar{m}^\ell_T(\theta_0, g) > c\right) = 0,$$

which then implies the second convergence in Lemma 3(b) because we have $[\bar{m}^\ell_T(\theta_0, g)]_+ = \max\{0, \bar{m}^\ell_T(\theta_0, g)\}$. □

C.3 Proof of Asymptotic Normality

To prove Theorem 2, we first give an auxiliary theorem that shows the convergence rate of $\widehat{\theta}_T$.

Theorem 3. *Suppose that either Assumption 1 holds and* $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$ *is bounded, or Assumption 2 holds and* $\sup_{t=1,\ldots,T} J_t$ *is bounded. Also suppose that Assumptions 3-9 hold. Then we have* $\widehat{\theta}_T^s - \theta_0^s = O_p(T^{-1/2})$.

Theorem 3 is proved using the following three lemmas. Theorem 3 and two of the lemmas together imply Theorem 2 as we explain immediately below. We give the proofs of Theorem 3 and the three lemmas in turn following the proof of Theorem 2.

Lemma 4. *Suppose that either Assumption 1 holds and* $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$ *is bounded, or Assumption 2 holds and* $\sup_{t=1,\ldots,T} J_t$ *is bounded. Also suppose that Assumptions 3-9 hold. Then we have for any sequence* θ_T *such that* $\theta_T^s - \theta_0^s = O_p(T^{-1/2})$, $\widehat{Q}_T(\theta_T) - \widehat{Q}_{0,T}(\theta_T) = o_p(T^{-1})$.

Lemma 5. *Suppose that either Assumption 1 holds and* $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$ *is bounded, or Assumption 2 holds and* $\sup_{t=1,\ldots,T} J_t$ *is bounded. Also suppose that Assumptions 3-9 hold. Then we have*

(a) *for an open ball* $B_c(\theta_0^s)$ *of radius* $c > 0$ *around* θ_0^s*, we have that*

$$\sup_{\theta\in\Theta:\theta^s\in B_c(\theta_0^s)} \left| \sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)} \right| = o_p(T^{-1/2}), \text{ and}$$

(b) $\widehat{Q}_T^*(\theta_0) = O_p(T^{-1})$.

Lemma 6. *Suppose that either Assumption 1 holds and* $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$ *is bounded, or Assumption 2 holds and* $\sup_{t=1,\ldots,T} J_t$ *is bounded. Also suppose that Assumptions 3-8 hold. For any sequence of random vectors* θ_T *such that* $\|\theta_T^s - \theta_0^s\| \to_p 0$*, we have*

(a) $\widehat{Q}_T^*(\theta_T) - \widehat{Q}_T^*(\theta_0) = (\theta_T^s - \theta_0^s)'\widehat{\Upsilon}_T(\theta_T^s - \theta_0^s) + 2W_T'(\theta_T^s - \theta_0^s) + o_p(1)\|\theta_T^s - \theta_0^s\|^2$*, where*

$$\widehat{\Upsilon}_T = \sum_{g\in\mathcal{G}_0} \mu(g)\widehat{\Gamma}_T(g)\widehat{\Gamma}_T(g)'$$

$$W_T = \sum_{g\in\mathcal{G}_0} \mu(g)\bar{m}_T(\theta_0, g)\widehat{\Gamma}_T(g)$$

$$\widehat{\Gamma}_T(g) = (T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t} g(z_{jt})\partial m_{jt}(\lambda_0), \text{ and}$$

(b) $\widehat{\Upsilon}_T \to_p \Upsilon$ *and* $T^{1/2}W_T \to_d N(0, V)$.

Proof of Theorem 2. We use Theorem 2 of **?** to prove the theorem. By Theorem 2 of **?**, the conclusion of our Theorem 2 holds under two conditions:

(i) $\|\widehat{\theta}^s_T - \theta^s_0\| = O_p(T^{-1/2})$,

(ii) uniformly over θ^s in a $O_p(T^{-1/2})$ neighborhood of θ^s_0, $\widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_0) = (\theta^s - \theta^s_0)'\Upsilon(\theta^s - \theta^s_0) + 2T^{-1/2}B'_T(\theta^s - \theta^s_0) + o_p(T^{-1})$ for a random vector B_T such that $B_T \to_d N(0, V)$.

Condition (i) is implied by Theorem 3. To establish condition (ii), consider the derivation: for any sequence θ_T such that $\theta^s_T - \theta^s_0 = O_p(T^{-1/2})$,

$$\begin{aligned}\widehat{Q}_T(\theta_T) - \widehat{Q}_T(\theta_0) &= [\widehat{Q}_T(\theta_T) - \widehat{Q}_{0,T}(\theta_T)] + [\widehat{Q}_{0,T}(\theta_T) - \widehat{Q}^*_T(\theta_T)] + \\ &\quad [\widehat{Q}^*_T(\theta_T) - \widehat{Q}^*_T(\theta_0)] + [\widehat{Q}^*_T(\theta_0) - \widehat{Q}_{0,T}(\theta_0)] + [\widehat{Q}_{0,T}(\theta_0) - \widehat{Q}_T(\theta_0)] \\ &= o_p(T^{-1}) + [\widehat{Q}_{0,T}(\theta_T) - \widehat{Q}^*_T(\theta_T)] + \\ &\quad [\widehat{Q}^*_T(\theta_T) - \widehat{Q}^*_T(\theta_0)] + [\widehat{Q}^*_T(\theta_0) - \widehat{Q}_{0,T}(\theta_0)] + o_p(T^{-1}), \end{aligned} \tag{74}$$

where the second equality holds by Lemma 4. For the summand $[\widehat{Q}_{0,T}(\theta_T) - \widehat{Q}^*_T(\theta_T)]$, consider the derivation:

$$\begin{aligned}\widehat{Q}_{0,T}(\theta_T) - \widehat{Q}^*_T(\theta_T) &= \left(\sqrt{\widehat{Q}_{0,T}(\theta_T)} - \sqrt{\widehat{Q}^*_T(\theta_T)}\right)^2 + 2\left(\sqrt{\widehat{Q}_{0,T}(\theta_T)} - \sqrt{\widehat{Q}^*_T(\theta_T)}\right)\left(\sqrt{\widehat{Q}^*_T(\theta_T)}\right) \\ &= o_p(T^{-1}) + o_p(T^{-1/2})\sqrt{\widehat{Q}^*_T(\theta_T) - \widehat{Q}^*_T(\theta_0) + \widehat{Q}^*_T(\theta_0)} \\ &= o_p(T^{-1}) + o_p(T^{-1/2})\sqrt{\widehat{Q}^*_T(\theta_T) - \widehat{Q}^*_T(\theta_0) + O_p(T^{-1})} \\ &= o_p(T^{-1}) + o_p(T^{-1/2})\sqrt{O_p(T^{-1}) + O_p(T^{-1})} \\ &= o_p(T^{-1}), \end{aligned} \tag{75}$$

where the second equality holds by Lemma 5(a), the third equality holds by Lemma 5(b), and the fourth equality holds by Lemma 6(a)-(b). Similar arguments show that the summand $[\widehat{Q}_{0,T}(\theta_0) - \widehat{Q}^*_T(\theta_0)] = o_p(T^{-1})$. Therefore,

$$\widehat{Q}_T(\theta) - \widehat{Q}_T(\theta_0) = o_p(T^{-1}) + \widehat{Q}^*_T(\theta_T) - \widehat{Q}^*_T(\theta_0) \tag{76}$$

This combined with Lemma 6(a)-(b) shows the condition (ii) where $B_T = T^{1/2}W_T$. This concludes the proof of Theorem 2. □

Proof of Theorem 3. We prove Theorem 3 using Lemmas 4-6. The three lemmas imply that

$$(eig_{min}(\Upsilon) + o_p(1))\|\widehat{\theta}^s_T - \theta^s_0\|^2 + O_p(T^{-1/2})\|\widehat{\theta}^s_T - \theta^s_0\|$$

$$
\begin{aligned}
&\leq \widehat{Q}_T^*(\widehat{\theta}_T) - \widehat{Q}_T^*(\theta_0) \\
&\leq (\sqrt{\widehat{Q}_{0,T}(\widehat{\theta}_T)} + o_p(T^{-1/2}))^2 - \widehat{Q}_T^*(\theta_0) \\
&\leq 2\widehat{Q}_{0,T}(\widehat{\theta}_T) + o_p(T^{-1}) - \widehat{Q}_T^*(\theta_0) \\
&\leq 2\widehat{Q}_T(\widehat{\theta}_T) + o_p(T^{-1}) - \widehat{Q}_T^*(\theta_0) \\
&\leq 2\widehat{Q}_T(\theta_0) + o_p(T^{-1}) - \widehat{Q}_T^*(\theta_0) \\
&\leq 2(\widehat{Q}_T(\theta_0) - \widehat{Q}_{0,T}(\theta_0)) + 2\widehat{Q}_{0,T}(\theta_0) + o_p(T^{-1}) \\
&= O_p(T^{-1}), \qquad (77)
\end{aligned}
$$

where $eig_{\min}(\Upsilon)$ is the smallest eigenvalue of Υ, the first inequality holds by Lemma 6(a)-(b), the second inequality holds by Lemma 5(a), the third inequality holds by the algebraic inequality $(a+b)^2 \leq 2a^2 + 2b^2$, the fourth inequality holds because $\widehat{Q}_{0,T}(\cdot)$ and $\widehat{Q}_T(\cdot)$ are defined to be exactly the same, both being weighted sums of nonnegative terms, except that the former sums over fewer terms, the fifth inequality holds because $\widehat{\theta}_T$ is the minimizer of $\widehat{Q}_T(\cdot)$, the sixth inequality holds because $\widehat{Q}_T^*(\theta_0) \geq 0$, and the equality holds by Lemmas 4 and 5(a)-(b). Let ζ be an arbitrary positive number, we next show that we can find a constant M_1 large enough so that

$$
\limsup_{T\to\infty} \Pr\left(T^{1/2}\|\widehat{\theta}_T^s - \theta_0^s\| > M_1\right) < \zeta. \qquad (78)
$$

This shows that $\|\widehat{\theta}_T^s - \theta_0^s\| = O_p(T^{-1/2})$. To show (78), consider that

$$
\begin{aligned}
&\Pr\left(T^{1/2}\|\widehat{\theta}_T^s - \theta_0^s\| > M_1\right) \\
&\leq \Pr\left(T^{1/2}\|\widehat{\theta}_T^s - \theta_0^s\| > M_1,\ o_p(1) \geq -eig_{\min}(\Upsilon)/2\right) + \Pr\left(o_p(1) < -eig_{\min}(\Upsilon)/2\right) \\
&\leq \Pr\left(T\|\widehat{\theta}_T^s - \theta_0^s\|^2(eig_{\min}(\Upsilon) + o_p(1)) > \frac{eig_{\min}(\Upsilon)M_1^2}{2},\ T^{1/2}\|\widehat{\theta}_T^s - \theta_0^s\| > M_1\right) + o(1) \\
&\leq \Pr\left(T\|\widehat{\theta}_T^s - \theta_0^s\|^2(eig_{\min}(\Upsilon) + o_p(1)) > \frac{eig_{\min}(\Upsilon)M_1^2}{2},\ T^{1/2}\|\widehat{\theta}_T^s - \theta_0^s\| > M_1,\ O_p(1) \geq -M_2\right) \\
&\qquad + \Pr\left(O_p(1) < -M_2\right) + o(1) \\
&\leq \Pr\left(T\|\widehat{\theta}_T^s - \theta_0^s\|^2(eig_{\min}(\Upsilon) + o_p(1)) + O_p(T^{1/2})\|\widehat{\theta}_T^s - \theta_0^s\| > \frac{eig_{\min}(\Upsilon)M_1^2}{2} - M_1M_2\right) \\
&\qquad + \Pr\left(O_p(1) < -M_2\right) + o(1) \\
&\leq \Pr\left(O_p(1) > \frac{eig_{\min}(\Upsilon)M_1^2}{2} - M_1M_2\right) + \Pr\left(O_p(1) < -M_2\right) + o(1),
\end{aligned}
$$

where the last inequality holds by (77), and the different $O_p(1)$ terms appearing above are not necessarily the same ones. Fix M_2 at a value such that the limsup of the second term in the last line is less than $\zeta/2$. Note that $\frac{eig_{\min}(\Upsilon)M_1^2}{2} - M_1 M_2$ can be made arbitrarily large by increasing M_1 (by Assumption 8(d), $eig_{\min}(\Upsilon) > 0$). Thus, we can choose a M_1 large enough so that the limsup of the first term in the last line is also less than $\zeta/2$. Therefore, a large enough M_1 exists such that (78) holds. □

Proof of Lemma 4. Note that

$$\widehat{Q}_T(\theta_T) - \widehat{Q}_{0,T}(\theta_T) = \sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[\bar{m}_T^u(\theta_T, g)]_-^2 + \sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[\bar{m}_T^\ell(\theta_T, g)]_+^2.$$

Thus, it suffices to show that

$$\sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[\bar{m}_T^u(\theta_T, g)]_-^2 = o_p(T^{-1}), \text{ and} \tag{79}$$

$$\sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[\bar{m}_T^\ell(\theta_T, g)]_+^2 = o_p(T^{-1}). \tag{80}$$

We separate the two cases, one where Assumption 1 is satisfied and the other where Assumption 2 is satisfied.

Case 1: Assumption 1 is satisfied. In this case, arguments for (79) and (80) are analogous. Thus, we give the detailed proof for (79) only. First consider that

$$\sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[\bar{m}_T^u(\theta_T, g)]_-^2 \leq \sum_{g \in \mathcal{G} \setminus \mathcal{G}_0} \mu(g)[A_T(g) + B_T(g) + C_T(g)]_-^2,$$

where

$$A_T(g) = \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\delta_{jt}(\pi_t, \lambda_T) - x_{jt}'\beta_T) g(z_{jt})$$

$$B_T(g) = \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\check{\delta}_{jt}(\tilde{s}_t, \lambda_T) - \check{\delta}_{jt}(\pi_t, \lambda_T)) g(z_{jt})$$

$$C_T(g) = \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\log(s_{jt} + \underline{\iota}_u/n_t) - \log(\pi_{jt})) g(z_{jt}). \tag{81}$$

The inequality holds because $\iota_u \geq \underline{\iota}_u$ (Assumption 1(b)). For $A_T(g)$, consider that

$$A_T(g) = \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} \xi_{jt} g(z_{jt}) + \frac{1}{T\bar{J}_T} \sum_{t=1}^{T} \sum_{j=1}^{J_t} (\delta_{jt}(\pi_t, \lambda_T) - \delta_{jt}(\pi_t, \lambda_0)) g(z_{jt})$$

$$- \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t} x_{jt}(\beta_T - \beta_0)g(z_{jt}).$$

Equation (64) in the proof of Lemma 3 implies that

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\xi_{jt}g(z_{jt})\right| = O_p(T^{-1/2}). \tag{82}$$

Also,

$$\begin{aligned}
&\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\pi_t,\lambda_T) - \delta_{jt}(\pi_t,\lambda_0))g(z_{jt})\right| \\
&= \sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\frac{\partial\delta_{jt}(\pi_t,\tilde{\lambda}_T)}{\partial\lambda'}(\lambda_T - \lambda_0)g(z_{jt})\right| \\
&\leq \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\|\frac{\partial\delta_{jt}(\pi_t,\tilde{\lambda}_T)}{\partial\lambda'}\right\|\|\lambda_T - \lambda_0\| \\
&= O_p(1)\|\lambda_T - \lambda_0\|,
\end{aligned} \tag{83}$$

where the first equality holds by a mean-value expansion for $\tilde{\lambda}_T$ lying on the line segment connecting λ_T and λ_0, the inequality holds because $g(z_{jt}) \in (0,1)$, the first equality holds by Assumption 8(c) and the condition that $\lambda_T \to \lambda_0$ given in the lemma. Moreover,

$$\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t} x_{jt}^s(\beta_T^s - \beta_0^s)g(z_{jt}) \leq \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\|x_{jt}^s\right\|\|\beta_T^s - \beta_0^s\| = O_p(1)\left\|\beta_T^s - \beta_0^s\right\|, \tag{84}$$

where the equality holds by Assumption 8(c).

Therefore, combining (82), (83), (84) and $\|\theta_T^s - \theta_0^s\| = O_p(T^{-1/2})$, we have

$$\sup_{g\in\mathcal{G}}|A_T(g)| = O_p(\log(T)T^{-1/2}). \tag{85}$$

Now consider $B_T(g)$. Let $B_T^0(g) = \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t,\lambda_0) - \check{\delta}_{jt}(\pi_t,\lambda_0))g(z_{jt})$. Consider that

$$\sup_{g\in\mathcal{G}}|B_T(g) - B_T^0(g)| \leq \sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\pi_t,\lambda_T) - \delta_{jt}(\pi_t,\lambda_0))g(z_{jt})\right|$$

$$+\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_T)-\delta_{jt}(\tilde{s}_t,\lambda_0))g(z_{jt})\right|.$$

The first summand is less than or equal to $O_p(1)\|\lambda_T-\lambda_0\|$ by (83). The second summand is also less than or equal to $O_p(1)\|\lambda_T-\lambda_0\|$ due to the same arguments as those for (83) and the convergence $\sup_{t=1,\dots,T}\|s_t-\pi_t\|_f\to_p 0$ implied by (47) and Assumption 4(g). Those combined with $\|\theta_T^s-\theta_0^s\|=O_p(T^{-1/2})$ shows that:

$$\sup_{g\in\mathcal{G}}|B_T(g)-B_T^0(g)|=O_p(T^{-1/2}). \tag{86}$$

For $B_T^0(g)$, consider that

$$\begin{aligned}
B_T^0(g)&=\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t,\lambda_0)-\check{\delta}_{jt}(\pi_t,\lambda_0))g(z_{jt})\\
&=\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}(\tilde{s}_t-s_t)\\
&+\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}(s_t-\pi_t)\\
&+\frac{1}{2T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})(\tilde{s}_t-\pi_t)'\frac{\partial^2\check{\delta}_{jt}(\tilde{\pi}_t,\lambda_0))}{\partial\pi\partial\pi'}(\tilde{s}_t-\pi_t),
\end{aligned} \tag{87}$$

where $\tilde{\pi}_t$ is a point on the line segment connecting $\tilde{s}_t$ and π_t. For the first summand, note that, by the Cauchy-Schwartz inequality and $g(z)\in[0,1]$, its absolute value is less than or equal to

$$\left(\sup_{t=1,\dots,T}n_t\|\tilde{s}_t-s_t\|_f\right)\left(\frac{1}{T\bar{J}_T\underline{n}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\|\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}\right\|\right)=O_p(1)\underline{n}_T^{-1}O_p(\sqrt{J_T^{\max}})=o_p(T^{-1/2}),$$

where the first equality holds by Assumption 4(f),

$$E\left(\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\left\|\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}\right\|\right)\le\sup_{j,t}E\left\|\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}\right\|=O(\sqrt{J_T^{\max}})$$

(by Assumption 9(b)), and Markov's inequality, and the second equality holds by Assumption

4(g). For the second summand of (87), we can apply Lemma 7 and get

$$
\begin{aligned}
&E\left[\sup_{g\in\mathcal{G}}\left(\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}(s_t-\pi_t)\right)^2\right]\\
&\leq\frac{C\sum_{t=1}^{T}J_t^2}{T^2\bar{J}_T^2}\max_{j,t}E\left(\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}(s_t-\pi_t)\right)^2\\
&=\frac{C\sum_{t=1}^{T}J_t^2}{T^2\bar{J}_T^2}\max_{j,t}E\left(\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}\frac{diag(\pi_t)-\pi_t\pi_t'}{n_t}\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi}\right)\\
&\leq\frac{C\sum_{t=1}^{T}J_t^2}{T^2\bar{J}_T^2\underline{n}_T}\max_{j,t}E\left(\left\|\frac{\partial\check{\delta}_{jt}(\pi_t,\lambda_0))}{\partial\pi'}\right\|^2\right)\\
&=O(\underline{n}_T^{-1}T^{-1}J_T^{\max})\\
&=o(T^{-1}),
\end{aligned}
$$

where the first equality holds by $E[(s_t-\pi_t)(s_t-\pi_t)'|\pi_t,n_t]=\frac{diag(\pi_t)-\pi_t\pi_t'}{n_t}$ which holds under Assumption 4(b), the second inequality holds because $diag(\pi_t)-\pi_t\pi_t'$ is positive semi-definite and its largest eigenvalue of does not exceed the highest π_{jt} which does not exceed 1 and because $n_t\geq\underline{n}_T$ for all $t=1,\ldots,T$, the second equality holds by Assumption 9(b) and the boundedness of $\sum_{t=1}^{T}J_t^2/(T\bar{J}_T^2)$, and the last equality holds by Assumption 4(g). Therefore, the Markov inequality applies and shows that the second summand of (87) is $o_p(T^{-1/2})$ uniformly over $g\in\mathcal{G}$. For the third summand of (87), consider that

$$
\begin{aligned}
&\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})(\tilde{s}_t-\pi_t)'\frac{\partial^2\check{\delta}_{jt}(\tilde{\pi}_t,\lambda_0))}{\partial\pi\partial\pi'}(\tilde{s}_t-\pi_t)\right|\\
&\leq_{w.p.a.1.}\sup_{j,t}\sup_{\pi:\|\pi-\pi_t\|\leq c}\left\|\frac{\partial^2\check{\delta}_{jt}(\pi,\lambda_0))}{\partial\pi\partial\pi'}\right\|T^{-1}\sum_{t=1}^{T}(\tilde{s}_t-\pi_t)'(\tilde{s}_t-\pi_t)\\
&\leq O_p(J_T^{\max})2\left[T^{-1}\sum_{t=1}^{T}\|\tilde{s}_t-s_t\|^2+T^{-1}\sum_{t=1}^{T}\|s_t-\pi_t\|^2\right]\\
&=O_p(J_T^{\max})O_p(\underline{n}_T^{-1})+O_p(J_T^{\max})T^{-1}\sum_{t=1}^{T}\|s_t-\pi_t\|^2\\
&=O_p(J_T^{\max})O_p(\underline{n}_T^{-1})+O_p(J_T^{\max})O_p(\underline{n}_T^{-1})\\
&=o_p(T^{-1/2}),
\end{aligned}
$$

where the first inequality holds because $\sup_t\|\tilde{s}_t-\pi_t\|\leq\sup_t\|\tilde{s}_t-\pi_t\|_f\leq c$ w.p.a.1. by Assumption 4(f,g) and equation (47) and also because $g(z)\in[0,1]$, the second inequality holds by Assumption 9(b), the first equality holds by Assumption 4(f), the second equality holds by Markov's inequality and $E\|s_t - \pi_t\|^2 = E\sum_{j=1}^{J_t} \pi_{jt}(1-\pi_{jt})/n_t \leq \underline{n}_T^{-1}$, and the last equality holds by Assumption 4(g). Combining the arguments for all the three summands in (87), we have

$$\sup_{g\in\mathcal{G}} \left|B_T^0(g)\right| = o_p(T^{-1/2}). \tag{88}$$

This and (86) together imply that

$$\sup_{g\in\mathcal{G}} |B_T(g)| = o_p(T^{-1/2}). \tag{89}$$

Next consider $C_T(g)$. Using the moment bound derived in (65) in the proof of Theorem 3 and the Markov inequality, we can derive

$$\sup_{g\in\mathcal{G}} |C_T(g) - E[C_T(g)]| = O_p\left(\frac{\log \overline{n}_T}{T^{1/2}}\right) = O_p\left(\frac{\log T}{T^{1/2}}\right), \tag{90}$$

where the second equality holds by $\overline{n}_T T^{-2} \to_p 0$ (Assumption 8(e)).

Let $r_T(g)$ denote $A_T(g)+B_T(g)+C_T(g)-E[C_T(g)]$. Then $\bar{m}_T^u(\theta_T, g) \geq r_T(g)+E[C_T(g)]$. And by equations (85), (89), and (90), we have

$$\sup_{g\in\mathcal{G}} |r_T(g)| = O_p(T^{-1/2}\log T). \tag{91}$$

For a sequence c_T such that $T^{-1/2}\log T = o(c_T)$, consider:

$$\begin{aligned}
\sum_{g\in\mathcal{G}\backslash\mathcal{G}_0 : E[C_T(g)]>c_T} \mu(g)[\bar{m}_T^u(\theta_T, g)]_-^2 &\leq \sum_{g\in\mathcal{G}\backslash\mathcal{G}_0 : E[C_T(g)]>c_T} \mu(g)[r_T(g)+c_T]_-^2 \\
&\leq \sup_{g\in\mathcal{G}} [r_T(g)+c_T]_-^2 \\
&= [o_p(c_T)+c_T]_-^2 \\
&=_{w.p.a.1} 0,
\end{aligned} \tag{92}$$

where the first inequality holds because $[\cdot]_-^2$ is nonincreasing, the second inequality holds because $\mu(g)$ is a probability mass function, the first equality holds by (91). Thus, the expression $\sum_{g\in\mathcal{G}\backslash\mathcal{G}_0 : E[C_T(g)]>c_T} \mu(g)[\bar{m}_T^u(\theta_T, g)]_-^2$ converges in probability to zero at arbitrary rate. Further restrict c_T so that $c_T = o((\log T)^{-2/\eta})$. This is possible because for any finite

$\eta > 0$, $\log(T)^{1+2/\eta} = o(T^{1/2})$. Also consider

$$\begin{aligned}
\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:E[C_T(g)]\le c_T} \mu(g)[\bar{m}_T^u(\theta_T,g)]_-^2 &\le \sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:E[C_T(g)]\le c_T} \mu(g)[r_T(g)]_-^2 \\
&\le \sup_{g\in\mathcal{G}} |r_T(g)|^2 \sum_{g\in\mathcal{G}\setminus\mathcal{G}_0:E[C_T(g)]\le c_T} \mu(g) \\
&= O_p\left(T^{-1}(\log T)^2\right) c_T^{\eta} \\
&= o_p(T^{-1}), \qquad (93)
\end{aligned}$$

where the first inequality holds because $\bar{m}_T^u(\theta_T,g) = r_T(g) + E[C_T(g)]$ and

$$E[C_T(g)] = (T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t} E[\log(s_{jt} + \underline{\iota}_u/n_t) - \log(\pi_{jt})g(z_{jt})] \ge 0$$

by the definition of $\underline{\iota}_u$, and the first equality holds by the first part of Assumption 9(a). Therefore, combining (92) and (93), we have

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0} \mu(g)[\bar{m}_T^u(\theta_T,g)]_-^2 = o_p(T^{-1}). \qquad (94)$$

Case 2: Assumption 2 is satisfied. We prove (79) first. Observe that

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0} \mu(g)[\bar{m}_T^u(\theta_T,g)]_-^2 = \sum_{g\in\mathcal{G}\setminus\mathcal{G}_0} \mu(g)[A_T(g) + \Delta_T(g) + S_T(g)]_-^2,$$

where

$$\begin{aligned}
A_T(g) &= \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\pi_t,\lambda_T) - x_{jt}'\beta_T)g(z_{jt}) \\
\Delta_T(g) &= \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_T) - \delta_{jt}(\pi_t,\lambda_T))g(z_{jt}) \\
S_T(g) &= \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt} + \iota_u/n_t) - \log(\tilde{s}_{jt}))g(z_{jt}). \qquad (95)
\end{aligned}$$

The same arguments showing (85) in Case 1 still applies in Case 2 since neither Assumption 1 or Assumption 2 is involved. Thus, (85) holds. For $\Delta_T(g)$, the same arguments as those for (86) shows that

$$\sup_{g\in\mathcal{G}\setminus\mathcal{G}_0}\left|\Delta_T(g)-\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})\right|=O_p(1)\|\lambda_T-\lambda_0\|. \tag{96}$$

Equation (69) in Case 2 of the proof of Theorem 3 shows that

$$\begin{aligned}&E\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})-E[(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})]\right|^2\\=&O\left(\frac{\log(\overline{n}_T)^2}{T}\right).\end{aligned}$$

Thus, by the Markov inequality,

$$\begin{aligned}&\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})-E[(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})]\right|\\=&O_p\left(\frac{\log(\overline{n}_T)}{T^{1/2}}\right)\\=&O_p(\log(T)T^{-1/2}).\end{aligned} \tag{97}$$

where the second equality holds by $\overline{n}_T T^{-2}\to_p 0$ (Assumption 8(e)). By Assumption 2(b), $E[(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})]\geq 0$. This combined with (96), (97), and $\|\widehat{\theta}_T^s-\theta_0^s\|=O_p(T^{-1/2})$ implies that

$$\inf_{g\in\mathcal{G}}\Delta_T(g)\geq O_p((\log(T)T^{-1/2}). \tag{98}$$

For $S_T(g)$, note that

$$\begin{aligned}S_T(g)&\geq\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(s_{jt}+\iota_u/n_t)^{-1}((s_{jt}+\iota_u/n_t)-(\tilde{s}_{jt}))g(z_{jt})\\&=(\iota_u-1)\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\frac{g(z_{jt})}{n_ts_{jt}+\iota_u}.\end{aligned}$$

Applying Lemma 7 and using the fact that $E[(n_ts_{jt}+\iota_u)^{-2}]\leq\iota_u^{-2}$, we have

$$E\sup_{g\in\mathcal{G}}\left(\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\frac{g(z_{jt})}{n_ts_{jt}+\iota_u}-E\left[\frac{g(z_{jt})}{n_ts_{jt}+\iota_u}\right]\right)^2=O(T^{-1}).$$

Then by the Markov inequality we have

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}\frac{g(z_{jt})}{n_t s_{jt}+\iota_u}-E\left[\frac{g(z_{jt})}{n_t s_{jt}+\iota_u}\right]\right|=O_p(T^{-1/2}).$$

Thus we have

$$S_T(g)\geq O_p(T^{-1/2})+\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}E\left[\frac{g(z_{jt})}{n_t s_{jt}+\iota_u}\right]. \tag{99}$$

Using (85), (98), (99), and the third part of Assumption 9(a), we can apply the same arguments as those for (94) (from (92) to (94)) to conclude that (79) holds.

Finally we prove (80) for Case 2. Note that

$$\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0}\mu(g)[\bar{m}_T^{\ell}(\theta_T,g)]_+^2\leq\sum_{g\in\mathcal{G}\setminus\mathcal{G}_0}\mu(g)[A_T(g)+B_T(g)+C_T^{\ell}(g)]_+^2,$$

where

$$\begin{aligned}
A_T(g)&=\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\pi_t,\lambda_T)-x_{jt}'\beta_T)g(z_{jt})\\
B_T(g)&=\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\check{\delta}_{jt}(\tilde{s}_t,\lambda_T)-\check{\delta}_{jt}(\pi_t,\lambda_T))g(z_{jt})\\
C_T^{\ell}(g)&=\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(s_{jt}+\bar{\iota}_\ell/n_t)-\log(\pi_{jt}))g(z_{jt}).
\end{aligned} \tag{100}$$

The same arguments showing (85) in Case 1 still applies in Case 2 since neither Assumption 1 or Assumption 2 is involved. Thus, (85) holds. For $B_T(g)$, the same arguments for (86) in Case 1 still applies here as well. Thus, (86) holds, and we only need to study $B_T^0(g)$ to understand the behavior of $B_T(g)$. Note that

$$\begin{aligned}
B_T^0(g)=&\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})-E[(\delta_{jt}(\tilde{s}_t,\lambda_0)-\delta_{jt}(\pi_t,\lambda_0))g(z_{jt})]\\
&-\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}(\log(\tilde{s}_t)-\log(\pi_t))g(z_{jt})-E[(\log(\tilde{s}_t)-\log(\pi_t))g(z_{jt})]\\
&+\frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t}E[(\check{\delta}_{jt}(\tilde{s}_t,\lambda_0)-\check{\delta}_{jt}(\pi_t,\lambda_0))g(z_{jt})].
\end{aligned}$$

Equation (97) shows that the first summand is $O_p(\log(T)T^{-1/2})$ uniformly over $g\in\mathcal{G}$, Equation (65) and Markov inequality combined show that the second summand is $O_p(\log(T)T^{-1/2})$ uniformly over $g \in \mathcal{G}$. The third summand is non-positive by Assumption 2(a). Therefore

$$\sup_{g\in\mathcal{G}} B_T(g) \le O_p(\log(T)T^{-1/2}). \tag{101}$$

The same arguments as those for the second summand above shows that $\sup_{g\in\mathcal{G}} |C_T^\ell(g) - E[C_T^\ell(g)]| = O_p(\log(T)T^{-1/2})$. Using this, (85), (101), and the second part of Assumption 9(a), we can apply similar arguments as those for (94) (from (92) to (94)) to conclude that (80) holds. $\square$

Proof of Lemma 5. (a) By equation (60) in the proof of Theorem 1, we have

$$\begin{aligned}&\sup_{\theta\in\Theta:\theta^s\in B_c(\theta_0^s)} \left|\sqrt{\widehat{Q}_{0,T}(\theta)} - \sqrt{\widehat{Q}_T^*(\theta)}\right| \\ &\le \sqrt{\sup_{\theta\in\Theta:\theta^s\in B_c(\theta_0^s)} \sup_{g\in\mathcal{G}_0} |\bar{m}_T^u(\theta,g) - \bar{m}_T(\theta,g)|^2 + \sup_{\theta\in\Theta:\theta^s\in B_c(\theta_0^s)} \sup_{g\in\mathcal{G}_0} |\bar{m}_T^\ell(\theta,g) - \bar{m}_T(\theta,g)|^2}.\end{aligned} \tag{102}$$

Now note that

$$\begin{aligned}\bar{m}_T^u(\theta,g) - \bar{m}_T(\theta,g) &= (T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}(\hat{\delta}_{jt}^u(s_t,\lambda) - \delta_{jt}(\pi_t,\lambda))g(z_{jt}) \\ &= (T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}(\log(s_{jt} + \iota_u/n_t) - \log(s_{jt}))g(z_{jt}) \\ &+ (T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}\delta_{jt}(\tilde{s}_t,\lambda) - \delta_{jt}(\pi_t,\lambda))g(z_{jt}).\end{aligned} \tag{103}$$

For the first summand, consider that

$$\begin{aligned}\sup_{g\in\mathcal{G}_0}\left|(T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}(\log(s_{jt} + \iota_u/n_t) - \log(s_{jt}))g(z_{jt})\right| &\le (T\bar{J}_T)^{-1}\iota_u\sum_{t=1}^T\sum_{j=1}^{J_t}(s_{jt} + \tilde{\iota}/n_t)^{-1}n_t^{-1} \\ &\le \underline{n}_T^{-1}\iota_u \sup_{j,t:z_{jt}\in\mathcal{Z}_0} s_{jt}^{-1} \\ &= O_p(\underline{n}_T^{-1}) = o_p(T^{-1/2}),\end{aligned} \tag{104}$$

where the first inequality holds with $\tilde{\iota} \in [0, \iota_u]$ by mean-value expansion and $|g(z_{jt})| \le 1$, the second inequality holds by the definition of $\mathcal{G}_0$, the first equality holds because s_{jt} is bounded away from zero by Assumptions 3 and equation (46), and the last equality holds by Assumption 8(e). For the second summand in (103), we can apply the same arguments as those for (89) to show that this second summand is $o_p(T^{-1/2})$ with the following adjustment: (1) Replace $\mathcal{G}$ by $\mathcal{G}_0$, (2) replace $\check{\delta}_{jt}(\cdot,\cdot)$ with $\delta_{jt}(\cdot,\cdot)$ and (2) replace the Case 1 version of Assumption 9(b) by the Case 2 version. Therefore, we have

$$\sup_{\theta\in\Theta:\theta^s\in B_c(\theta_0^s)}\sup_{g\in\mathcal{G}_0}|\bar{m}_T^u(\theta,g)-\bar{m}_T(\theta,g)|=o_p(T^{-1/2}).$$

Analogous arguments can be used to show that $\sup_{\theta\in B_c(\theta_0)}\sup_{g\in\mathcal{G}_0}|\bar{m}_T^\ell(\theta,g)-\bar{m}_T(\theta,g)|=o_p(T^{-1/2})$. That concludes the proof. □

(b) Recall that $\widehat{Q}_T^*(\theta_0)=\sum_{g\in\mathcal{G}_0}\mu(g)(\bar{m}_T(\theta_0,g))^2$, and note that

$$\bar{m}_T(\theta_0,g)=\frac{1}{T\bar{J}_T}\sum_{T=1}^{T}\sum_{j=1}^{J_t}(\delta_{jt}(\pi_t,\lambda_0)-x_{jt}'\beta_0)g(z_{jt})=\frac{1}{T}\sum_{T=1}^{T}\sum_{j=1}^{J_t}\xi_{jt}g(z_{jt}).$$

Then by equation(64) in the proof of Theorem 3, we have

$$\sup_{g\in\mathcal{G}_0}|\bar{m}_T(\theta_0,g)|=O_p(T^{-1/2}). \tag{105}$$

This implies part (b).

Proof of Lemma 6. (a) First consider that, for $g\in\mathcal{G}_0$,

$$\begin{aligned}
&\bar{m}_T(\theta_T,g)-\bar{m}_T(\theta_0,g)\\
&=(T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})[\delta_{jt}(\pi_t,\lambda_T)-\delta_{jt}(\pi_t,\lambda_0)+x_{jt}^s(\beta_T^s-\beta_0^s)]\\
&=(T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})\partial m_{jt}(\lambda_0)'(\theta_T^s-\theta_0^s)+\\
&\qquad(T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})(\lambda_T-\lambda_0)'\frac{\partial^2\delta_{jt}(\pi_t,\tilde{\lambda})}{\partial\lambda\partial\lambda'}(\lambda_T-\lambda_0)/2\\
&=\widehat{\Gamma}_T(g)'(\theta_T^s-\theta_0^s)+(\lambda_T-\lambda_0)'D_T(g)(\lambda_T-\lambda_0),
\end{aligned}$$

where $\tilde{\lambda}$ is a point on the line segment connecting λ_T and λ_0, and

$$D_T(g)=(2T\bar{J}_T)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{J_t}g(z_{jt})\frac{\partial^2\delta_{jt}(\pi_t,\tilde{\lambda})}{\partial\lambda\partial\lambda'}.$$

Thus, we have

$$
\begin{aligned}
&\widehat{Q}_T^*(\theta_T) - \widehat{Q}_T^*(\theta_0) \\
&= \sum_{g\in\mathcal{G}_0} \mu(g)(\bar{m}_T(\theta_T, g) - \bar{m}_T(\theta_0, g))^2 + 2\sum_{g\in\mathcal{G}_0} \mu(g)\bar{m}_T(\theta_0, g)(\bar{m}_T(\theta_T, g) - \bar{m}_T(\theta_0, g)) \quad (106) \\
&= (\theta_T^s - \theta_0^s)\sum_{g\in\mathcal{G}_0} \mu(g)\widehat{\Gamma}_T(g)\widehat{\Gamma}_T(g)'(\theta_T^s - \theta_0^s) \\
&\qquad + 2\sum_{g\in\mathcal{G}_0} \mu(g)(\lambda_T - \lambda_0)'D_T(g)(\lambda_T - \lambda_0)\widehat{\Gamma}_T(g)'(\theta_T^s - \theta_0^s) \\
&\qquad + \sum_{g\in\mathcal{G}_0} \mu(g)\{(\lambda_T - \lambda_0)'D_T(g)(\lambda_T - \lambda_0)\}^2 \\
&\qquad + 2\sum_{g\in\mathcal{G}_0} \mu(g)\bar{m}_T(\theta_0, g)\widehat{\Gamma}_T(g)'(\theta_T^s - \theta_0^s) \\
&\qquad + 2\sum_{g\in\mathcal{G}_0} \mu(g)\bar{m}_T(\theta_0, g)(\lambda_T - \lambda_0)'D_T(g)(\lambda_T - \lambda_0). \quad (107)
\end{aligned}
$$

Since $\tilde{\lambda} \in B_c(\lambda_0)$ whenever $\lambda_T \in B_c(\lambda_0)$ (which holds with probability approaching one because $\|\lambda_T - \lambda_0\| \to_p 0$), we have for any $g \in \mathcal{G}_0$,

$$
\sup_{g\in\mathcal{G}_0} \|D_T(g)\| \le_{w.p.a.1} \frac{1}{T\bar{J}_T}\sum_{t=1}^{T}\sum_{j=1}^{J_t} \sup_{\lambda:\|\lambda-\lambda_0\|\le c} \left\| \frac{\partial^2 \delta_{jt}(\pi_t, \lambda)}{\partial\lambda\partial\lambda'} \right\| = O_p(1), \quad (108)
$$

where the first inequality holds because $0 \le g(z) \le 1$, and the equality holds by Markov's inequality and Assumption 8(c). This combined with $\|\theta_T^s - \theta_0^s\| = o_p(1)$ implies that

$$
\sum_{g\in\mathcal{G}_0} \mu(g)\{(\lambda_T - \lambda_0)'D_T(g)(\lambda_T - \lambda_0)\}^2 \le \sup_{g\in\mathcal{G}_0} \|D_T(g)\|^2 \|\theta_T^s - \theta_0^s\|^4 = o_p(1)\|\theta_T^s - \theta_0^s\|^2.
$$

Also, using Assumption 8(c) and the same arguments as those for (108), we can show that $\sup_{g\in\mathcal{G}_0} \|\widehat{\Gamma}_T(g)\| = O_p(1)$. This combined with (108) and $\|\theta_T^s - \theta_0^s\| = o_p(1)$ implies that

$$
\begin{aligned}
\left| \sum_{g\in\mathcal{G}_0} \mu(g)(\lambda_T - \lambda_0)'D_T(g)(\lambda_T - \lambda_0)\widehat{\Gamma}_T(g)'(\theta_T^s - \theta_0^s) \right| &\le \|\theta_T^s - \theta_0^s\|^3 \sup_{g\in\mathcal{G}_0} \|D_T(g)\|\|\widehat{\Gamma}_T(g)\| \\
&= o_p(1)\|\theta_T^s - \theta_0^s\|^2.
\end{aligned}
$$

Next apply Lemma 7 with $w_{jt} = \xi_{jt}$ and we get

$$
E\sup_{g\in\mathcal{G}_0} (\bar{m}_T(\theta_0, g))^2 = E\sup_{g\in\mathcal{G}_0} \left((T\bar{J}_T)^{-1} \sum_{t=1}^{T}\sum_{j=1}^{J_t} \xi_{jt} g(z_{jt}) \right)^2
$$

$$\leq \frac{C\sum_{t=1}^T J_t^2}{T^2\bar{J}_T^2}\sup_{j,t} E[\xi_{jt}^2 1(z_{jt}\in\mathcal{Z}_0)]$$
$$= O(T^{-1}),$$

where the second equality holds by Assumption 4(e) and the boundedness of $T^{-1}\sum_{t=1}^T J_t^2/\bar{J}_T^2$. Therefore,

$$\sup_{g\in\mathcal{G}_0}|\bar{m}_T(\theta_0,g)| = O_p(T^{-1/2}). \tag{109}$$

This combined with (108) implies that

$$\sum_{g\in\mathcal{G}_0}\mu(g)\bar{m}_T(\theta_0,g)(\lambda_T-\lambda_0)'D_T(g)(\lambda_T-\lambda_0) = O_p(T^{-1/2})\|\theta_T^s-\theta_0^s\|^2.$$

Therefore, part (a) holds.

(b) Apply Lemma 7 with w_{jt} being an element of the random vector $\partial m_{jt}(\lambda_0)$, do so for every element of $\partial m_{jt}(\lambda_0)$, and we get

$$E\sup_{g\in\mathcal{G}_0}\left\|\widehat{\Gamma}_T(g)-\Gamma_T(g)\right\|^2 \leq \frac{C\sum_{t=1}^T J_t^2}{T^2\bar{J}_T^2}\sup_{j,t}E[\|\partial m_{jt}(\lambda_0)\|^2 1(z_{jt}\in\mathcal{Z}_0)] = O(T^{-1}).$$

The equality is implied by Assumptions 8(c) and the boundedness of J_t. Thus, we have

$$\sup_{g\in\mathcal{G}_0}\left\|\widehat{\Gamma}_T(g)-\Gamma_T(g)\right\| = O_p(T^{-1/2}). \tag{110}$$

Assumption 8(c) also implies that

$$\sup_{g\in\mathcal{G}_0}\|\Gamma_T(g)\| \leq \sup_{g\in\mathcal{G}_0}(T\bar{J}_T)^{-1}\sum_{t=1}^T\sum_{j=1}^{J_t}E[\|\partial m_{jt}(\lambda_0)g(z_{jt})\|]$$
$$\leq \sup_{g\in\mathcal{G}_0}\sup_{j,t}E[\|\partial m_{jt}(\lambda_0)\|1(z_{jt}\in\mathcal{Z}_0)]$$
$$= O(1). \tag{111}$$

This and (110) together imply that

$$\widehat{\Upsilon}_T = \sum_{g\in\mathcal{G}_0}\mu(g)\widehat{\Gamma}_T(g)\widehat{\Gamma}_T(g)' = o_p(1) + \sum_{g\in\mathcal{G}_0}\mu(g)\Gamma_T(g)\Gamma_T(g)' \to_p \Upsilon,$$

where the convergence holds by Assumption 8(d).

For W_n, first consider the derivation

$$\left| T^{1/2} \sum_{g \in \mathcal{G}_0} \mu(g) \bar{m}_T(\theta_0, g)(\widehat{\Gamma}_T(g) - \Gamma_T(g)) \right| \leq \sup_{g \in \mathcal{G}_0} |\bar{m}_T(\theta_0, g)| \sup_{g \in \mathcal{G}_0} T^{1/2} \|\widehat{\Gamma}_T(g) - \Gamma_T(g)\|$$
$$= O_p(T^{-1/2}) = o_p(1),$$

by equations (109) and (110). Thus,

$$T^{1/2} W_n = o_p(1) + T^{1/2} \sum_{g \in \mathcal{G}_0} \mu(g) \bar{m}_T(\theta_0, g) \Gamma_T(g)$$
$$= o_p(1) + T^{-1/2} \sum_{t=1}^{T} v_t,$$

where $v_t = \bar{J}_T^{-1} \sum_{j=1}^{J_t} \left[\xi_{jt} \left(\sum_{g \in \mathcal{G}_0} \mu(g) g(z_{jt}) \Gamma_T(g) \right) \right]$. Observe that $\{v_t\}_{t=1}^T$ is independent across t by Assumption 4(d). Also consider the derivation:

$$E[v_t] = E \sum_{j=1}^{J_t} \left[E[\xi_{jt} | z_{jt}] \left(\sum_{g \in \mathcal{G}_0} \mu(g) g(z_{jt}) \Gamma_T(g) \right) \right] = 0$$
$$T^{-1} \sum_{t=1}^{T} E[v_t v_t']$$
$$= T^{-1} \sum_{t=1}^{T} \sum_{g, g^* \in \mathcal{G}_0} Cov \left(\bar{J}_T^{-1} \sum_{j=1}^{J_t} \xi_{jt} g(z_{jt}), \bar{J}_T^{-1} \sum_{j=1}^{J_t} \xi_{jt} g^*(z_{jt}) \right) \Gamma_T(g) \Gamma_T(g)' \mu(g) \mu(g^*) \to V,$$

where the second equality in the first lines holds by Assumptions 4(c), and the convergence holds by 8(d). Also for the c in Assumption 4(e),

$$E(\|v_t\|^{2+c}) \leq \sup_{j,t} E|\xi_{jt}|^{2+c} \sup_{g \in \mathcal{G}_0} \|\Gamma_T(g)\|^{2+c}$$
$$= O(1),$$

by Assumptions 4(d) and equation (111) above. Therefore, we can apply the Lindeberg central limit theorem and conclude $T^{-1/2} \sum_{t=1}^{T} v_t \to_d N(0, V)$. Therefore,

$$T^{1/2} W_n \to_d N(0, V).$$

□

C.4 Auxiliary Lemmas

In this subsection, we present two auxiliary lemmas. Lemma 7 establishes a maximal inequality for certain empirical processes indexed by g in a subset of $\mathcal{G}$. This is used above to establish the convergence rates of several empirical processes. Lemma 8 establishes a concentration inequality for the L_2 distance between a multinomial random vector and its expectation. This is used to derive a tighter tail bound for $\|s_t - \pi_t\|$ than that implied by Chernoff's inequality when J_t is large.

Lemma 7. *Let $\{z_{jt} : j = 1, \ldots, J_t, t = 1, \ldots, T\}_{T \geq 1}$ be an array of random vectors. Let $\mathcal{G}$ be the set of instrumental functions defined in (16). Let $\mathcal{Z}^*$ be a subset of supp(z_{jt}) and let $\mathcal{G}^*$ be a subset of $\mathcal{G}$ for which $g(z) = 0$ for all $z \notin \mathcal{Z}^*$for all $g \in \mathcal{G}^*$. Let $\{w_{jt} : j = 1, \ldots, J_t, t = 1, \ldots, T\}_{T\geq 1}$ be an array of random variables such that $E[w_{jt}^2 1(z_{jt} \in \mathcal{Z}^*)] \leq M_T$ for all j, t for some $M_T < \infty$. Let $w_t = (w_{1t}, \ldots, w_{J_t t})'$ and $z_t = (z_{1t}, \ldots, z_{J_t t})'$. Suppose that (w_t, z_t) is independent across t. Then*

$$E \sup_{g\in\mathcal{G}^*} \left(\sum_{t=1}^{T} \sum_{j=1}^{J_t} (w_{jt} g(z_{jt}) - E[w_{jt} g(z_{jt})]) \right)^2 \leq C M_T \sum_{t=1}^{T} J_t^2,$$

for some constant $C > 0$.

Proof. Recall that $J_T^{\max} = \max_{t=1,\ldots,T} J_t$. First observe that $\sum_{j=1}^{J_t} w_{jt} g(z_{jt})$ can be written as $f_t(g) := \sum_{j=1}^{J_T^{\max}} w_{jt} 1(j \leq J_t) g(z_{jt})$. Observe that the triangular array of random processes $\{g(z_{jt}) : g \in \mathcal{G}^* : t = 1, \ldots, T\}_{T\geq 1}$ is manageable with respect to the envelope $\mathbf{1}_T$ for all j in the sense of Pollard (1990) because $\mathcal{G}$ is the collection of indicator functions for a Vapnik-Cervonenkis class of sets. Then by parts (a) and (c) of Lemma E1 in Andrews and Shi (2013), we have that the triangular array $\{f_t(g) : g \in \mathcal{G}^* : t = 1, \ldots, T; T \geq 1\}$ is manageable with respect to the envelope function $F_T = (F_{T1}, \ldots, F_{TT})$ where $F_{Tt} = \sum_{j=1}^{J^{\max}} 1(j \leq J_t, z_{jt} \in \mathcal{Z}^*)|w_{jt}| \equiv \sum_{j=1}^{J_t} |w_{jt}| 1(z_{jt} \in \mathcal{Z}^*)$. Therefore, by the maximal inequality (7.10) in **?**, we have, for some constant $C > 0$,

$$\begin{aligned} E \sup_{g\in\mathcal{G}^*} \left| \sum_{t=1}^{T} \sum_{j=1}^{J_t} (w_{jt} g(z_{jt}) - E[w_{jt} g(z_{jt})]) \right|^2 &\leq C \sum_{t=1}^{T} E[(F_{Tt})^2] \\ &\leq C \sum_{t=1}^{T} J_t \sum_{j=1}^{J_t} E[w_{jt}^2 1(z_{jt} \in \mathcal{Z}^*)] \\ &\leq C M_T \sum_{t=1}^{T} J_t^2, \end{aligned} \tag{112}$$

proving the lemma. $\square$

The following lemma presents a concentration inequality for the L_2 distance from the mean for multinomial random vectors. The tail bound presented here does not depend on the length of the multinomial random vector, and thus can be applied for multinomial distributions with an arbitrarily large number of categories. The proof of the lemma uses Poissonization, a technique that **?** employs in his Lemma 3 to derive a concentration inequality for the L_1 distance from the mean for multinomial random vectors. Devroye's bound applies when the number of categories is smaller than a scalar multiple of the sample size.

Lemma 8. *Let $(X_1, \ldots, X_J)$ be a multinomial $(n, p_1, \ldots, p_J)$ random vector, where $p_1, \ldots, p_J$ are non-negative numbers that sum up to 1 and n is a positive integer. Then, for all $\varepsilon > 0$,*

$$\Pr\left(\sum_{j=1}^{J}(X_i - np_i)^2 > n^2\varepsilon^2\right) \leq \frac{128(3n^2+n)}{n^4\varepsilon^4}. \tag{113}$$

Proof. Let $U_1, U_2, \ldots$ be a sequence of independent and identically distributed $\{1, \ldots, J\}$-valued variables with probability mass given by $P(U_1 = j) = p_j, 1 \leq j \leq J$. Let N be a Poisson(n) random variable independent of $\{U_1, U_2, \ldots\}$. Let X_j be the number of occurrences of the value j among $U_1, \ldots, U_n$, and let $\tilde{X}_i$ be the number of occurrences of the value j among $U_1, \ldots, U_N$. It is clear that $X_1, \ldots, X_J$ is a multinomial $(n, p_1, \ldots, p_J)$ random vector, and that $\tilde{X}_1, \ldots, \tilde{X}_J$ are independent Poisson random variables with means $np_1, \ldots, np_J$. By the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\sum_{j=1}^{J}(X_j - np_j)^2 \leq 2\sum_{j=1}^{J}(X_j - \tilde{X}_j)^2 + 2\sum_{j=1}^{J}(\tilde{X}_j - np_j)^2. \tag{114}$$

Thus,

$$\begin{aligned}
&\Pr\left(\sum_{j=1}^{J}(X_j - np_j)^2 > n^2\varepsilon^2\right) \\
&\leq \Pr\left(\sum_{j=1}^{J}(X_j - \tilde{X}_j)^2 > n^2\varepsilon^2/4\right) + \Pr\left(\sum_{j=1}^{J}(\tilde{X}_j - np_j)^2 > n^2\varepsilon^2/4\right).
\end{aligned} \tag{115}$$

For $\sum_{j=1}^{J}(X_j - \tilde{X}_j)^2$ consider the derivation:

$$(X_j - \tilde{X}_j)^2 = \left(\sum_{i=n+1}^{N} 1\{U_i = j\}\right)^2 1\{n < N\} + \left(\sum_{i=N+1}^{n} 1\{U_i = j\}\right)^2 1\{n > N\}$$

$$= \left(\sum_{i=n+1}^{N} 1\{U_i = j\} + 2 \sum_{i \neq i', i, i' = n+1, \ldots, N} 1\{U_i = j\} 1\{U_{i'} = j\} \right) 1\{n < N\}$$

$$+ \left(\sum_{i=N+1}^{n} 1\{U_i = j\} + 2 \sum_{i \neq i', i, i' = N+1, \ldots, n} 1\{U_i = j\} 1\{U_{i'} = j\} \right) 1\{n > N\}. \quad (116)$$

Thus

$$\sum_{j=1}^{J} (X_j - \tilde{X}_j)^2 = \left(\left(\sum_{i=n+1}^{N} 1 \right) + \sum_{i \neq i' : i, i' = n+1, \ldots, N} 1\{U_i = U_{i'}\} \right) 1\{n < N\}$$

$$+ \left(\left(\sum_{i=N+1}^{n} 1 \right) + \sum_{i \neq i' : i, i' = N+1, \ldots, n} 1\{U_i = U_{i'}\} \right) 1\{n > N\}$$

$$\leq |N - n + (N - n)(N - n - 1)|$$

$$= (N - n)^2. \quad (117)$$

Therefore, using Markov's inequality, we have

$$\Pr \left(\sum_{j=1}^{J} (X_j - \tilde{X}_j)^2 > n^2 \varepsilon^2 / 4 \right) \leq \Pr(|N - n|^2 > n^2 \varepsilon^2 / 4)$$

$$\leq \frac{16 E[(N - n)^4]}{n^4 \varepsilon^4} = \frac{16(3n^2 + n)}{n^4 \varepsilon^4}. \quad (118)$$

where the equality holds by $N \sim Poisson(n)$. For $\sum_{j=1}^{J} (\tilde{X}_j - np_j)^2$ consider the derivation:

$$E \left[\left(\sum_{j=1}^{J} (\tilde{X}_j - np_j)^2 \right)^2 \right]$$

$$= \sum_{j=1}^{J} E[(\tilde{X}_j - np_j)^4] + \sum_{j \neq j' : j, j' = 1, \ldots, J} E[(\tilde{X}_j - np_j)^2] E[(\tilde{X}_{j'} - np_{j'})^2]$$

$$= \sum_{j=1}^{J} (3n^2 p_j^2 + np_j) + \sum_{j \neq j' : j, j' = 1, \ldots, J} n^2 p_j p_j'$$

$$= 2n^2 \sum_{j=1}^{J} p_j^2 + n + n^2 \leq 3n^2 + n. \quad (119)$$

Therefore,

$$\Pr\left(\sum_{j=1}^{J}(\tilde{X}_j - np_j)^2 > n^2\varepsilon^2/4\right) \le \frac{16(3n^2+n)}{n^4\varepsilon^4} \tag{120}$$

Equations (115), (118), and (120) together concludes the proof. □

D Random Coefficient Logit

In this section, we prove a lemma that establishes Assumption 1 for the random coefficient logit model.

Lemma 9. *Consider the random coefficient logit model in Example 4.2. Also assume that (i) w_{jt} is bounded, i.e. $||w_{jt}|| \le \overline{w}$; (ii) $\sup_{\lambda\in\Lambda}\sup_{\|w\|\le\overline{w}}\int \exp(2w'v)dF(v;\lambda) < \infty$, (iii) $\inf_{t=1,\dots,T}\inf_{\pi_t\in\Delta^0_{J_t}}\pi_{0t} \ge \underline{\varepsilon}_0 > 0$ for all T, and (iv) there exists $e_1 > 0$ and $0 < e_2 < \underline{\varepsilon}_0/2$ such that, the maximum eigenvalue of $\int \tilde{\pi}_t(v)\tilde{\pi}_t(v)'dF(v;\lambda)\begin{pmatrix}\tilde{\pi}_{1t}^{-1} & 0 & \dots & 0\\ 0 & \tilde{\pi}_{2t}^{-1} & \dots & 0\\ \dots & \dots & \ddots & 0\\ 0 & 0 & 0 & \tilde{\pi}_{J_t t}^{-1}\end{pmatrix}$ is less than $1-e_1$ for all $\lambda\in\Lambda$, and all $\tilde{\pi}_t\in\Delta^{e_2}_{J_t}$ for all $t=1,\dots,T$ and $T=1,2,3,\dots$, where*

$$\tilde{\pi}_{jt}(v) = \frac{\exp(w'_{jt}v + \delta_{jt}(\tilde{\pi}_t;\lambda))}{1+\sum_{k=1}^{J_t}\exp(w'_{kt}v + \delta_{kt}(\tilde{\pi}_t;\lambda))}.$$

Then Assumption 1(a) *is satisfied.*

Proof. Without loss of generality, consider the derivative with respect to $\pi_{\ell t}$. For $j = 1,\dots,J_t$, take partial derivative with respect to $\pi_{\ell t}$ on both sides of (24), and we get:

$$\begin{aligned}
&\frac{\partial\check{\delta}_{jt}(\tilde{\pi}_t;\lambda)}{\partial\pi_{\ell t}}\\
&=\int \frac{\exp(w'_{jt}v)\exp(\check{\delta}_{jt}(\tilde{\pi}_t;\lambda))}{\left(1+\sum_{k=1}^{J_t}\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)+w'_{kt}v)\tilde{\pi}_{kt}\right)^2}\\
&\cdot\left(\exp(\check{\delta}_{\ell t}(\tilde{\pi}_t;\lambda)+w'_{\ell t}v) + \sum_{k=1}^{J_t}\tilde{\pi}_{kt}\exp(w'_{kt}v)\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda))\frac{\partial\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)}{\partial\pi_{\ell t}}\right)dF(v;\lambda)\\
&=\tilde{\pi}_{\ell t}^{-1}\tilde{\pi}_{jt}^{-1}\int\tilde{\pi}_{jt}(v)\tilde{\pi}_{\ell t}(v)dF(v;\lambda) + \sum_{k=1}^{J_t}\left\{\left[\tilde{\pi}_{jt}^{-1}\int\tilde{\pi}_{jt}(v)\tilde{\pi}_{kt}(v)dF(v;\lambda)\right]\frac{\partial\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)}{\partial\pi_{\ell t}}\right\}.
\end{aligned}$$

Stacking the J_t equations in matrix form, we find that

$$H_t(\tilde{\pi}_t, \lambda)\frac{\partial \check{\delta}_t(\tilde{\pi}_t; \lambda)}{\partial \pi_{1t}} = b_{\ell t}(\tilde{\pi}_t; \lambda),$$

where

$$H_t(\tilde{\pi}_t, \lambda) = I - \int \tilde{\pi}_t(v)\tilde{\pi}_t(v)' dF(v; \lambda) \begin{pmatrix} \tilde{\pi}_{1t}^{-1} & 0 & \dots & 0 \\ 0 & \tilde{\pi}_{2t}^{-1} & \dots & 0 \\ \dots & \dots & \ddots & 0 \\ 0 & 0 & 0 & \tilde{\pi}_{J_t t}^{-1} \end{pmatrix},$$

and

$$b_{\ell t}(\tilde{\pi}_t; \lambda) = \begin{pmatrix} \tilde{\pi}_{\ell t}^{-1}\tilde{\pi}_{1t}^{-1} \int \tilde{\pi}_{\ell t}(v)\tilde{\pi}_{1t}(v) dF(v; \lambda) \\ \tilde{\pi}_{\ell t}^{-1}\tilde{\pi}_{2t}^{-1} \int \tilde{\pi}_{\ell t}(v)\tilde{\pi}_{2t}(v) dF(v; \lambda) \\ \vdots \\ \tilde{\pi}_{\ell t}^{-1}\tilde{\pi}_{J_t t}^{-1} \int \tilde{\pi}_{\ell t}(v)\tilde{\pi}_{J_t t}(v) dF(v; \lambda) \end{pmatrix}.$$

By condition (iv), we have that the eigenvalues of $H_t(\tilde{\pi}_t, \lambda)$ are positive and bounded away from zero for all t, all λ and all $\tilde{\pi}_t \in \Delta_{J_t}^{e_2}$. Next we show that the elements $b_{\ell t}(\tilde{\pi}_t; \lambda)$ is bounded uniformly over ℓ and t, which will then imply that

$$\sup_{t=1,\dots,T; T=1,2,\dots} \sup_{j,\ell=1,\dots,J_t} \sup_{\tilde{\pi}_t \in \Delta_{J_t}^{e_2}} \sup_{\lambda \in \Lambda} \left| \frac{\partial \check{\delta}_{jt}(\tilde{\pi}_t; \lambda)}{\partial \pi_{\ell t}} \right| \le M < \infty.$$

for some M. Consider the derivation

$$\begin{aligned} \check{\delta}_{jt}(\hat{\pi}_t; \lambda) - \check{\delta}_{jt}(\pi_t; \lambda) &= \frac{\partial \check{\delta}_{jt}(\tilde{\pi}_t; \lambda)}{\partial \pi_t'}(\hat{\pi}_t - \pi_t) \\ &\le \|\hat{\pi}_t - \pi_t\| \left\| \frac{\partial \check{\delta}_{jt}(\tilde{\pi}_t; \lambda)}{\partial \pi_t'} \right\| \\ &\le \sqrt{J_t} M \|\hat{\pi}_t - \pi_t\| \\ &\le \sqrt{J_t} M \|\hat{\pi}_t - \pi_t\|_f. \end{aligned} \tag{121}$$

Thus Assumption 1(a) holds.

To show that $b_{\ell t}(\tilde{\pi}_t; \lambda)$ is uniformly bounded, we first show that $\check{\delta}_{jt}(\tilde{\pi}_t; \lambda)$ is uniformly bounded. Without loss of generality, consider $\check{\delta}_{1t}(\tilde{\pi}_t; \lambda)$:

$$\begin{aligned} \check{\delta}_{1t}(\tilde{\pi}_t; \lambda) &= -\log \int \frac{\exp(w_{jt}' v)}{1 + \sum_{k=1}^{J_t} \exp(\check{\delta}_{kt}(\tilde{\pi}_t; \lambda) + w_{kt}' v)\tilde{\pi}_{kt}} dF(v; \lambda) \\ &\ge -\log \int \exp(w_{jt}' v) dF(v; \lambda) \end{aligned}$$

$$\geq -\log \sup_{\lambda\in\Lambda} \sup_{\|w\|\leq\overline{w}} \int \exp(w'v)dF(v;\lambda),$$

where the second inequality holds by condition (i). Then by condition (ii), we have $\inf_{t,\lambda,\tilde{\pi}_t} \check{\delta}_{1t}(\tilde{\pi}_t;\lambda) > -\infty$. To show that $\sup_{t,\lambda,\tilde{\pi}_t} \check{\delta}_{1t}(\tilde{\pi}_t;\lambda) < \infty$, consider the outside share:

$$\tilde{\pi}_{0t} = \int \frac{1}{1+\sum_{k=1}^{J_t}\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)+w_{kt}'v)\tilde{\pi}_{kt}} dF(v;\lambda).$$

By $|\tilde{\pi}_{0t}-\pi_{0t}| \leq \|\tilde{\pi}_t-\pi_t\| < e_2 < \underline{\varepsilon}_0/2$ and $\pi_{0t} \geq \underline{\varepsilon}_0$, we have $\tilde{\pi}_{0t} \geq \underline{\varepsilon}_0/2$. Then there must exists $\overline{v}$ large enough such that $\int_{\|v\|\leq\overline{v}} \frac{1}{1+\sum_{k=1}^{J_t}\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)+w_{kt}'v)\tilde{\pi}_{kt}} dF(v;\lambda) \geq \underline{\varepsilon}_0/4$. Then

$$\begin{aligned}
\check{\delta}_{1t}(\tilde{\pi};\lambda) &\leq -\log \int_{\|v\|\leq\overline{v}} \frac{\exp(w_{jt}'v)}{1+\sum_{k=1}^{J_t}\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)+w_{kt}'v)\tilde{\pi}_{kt}} dF(v;\lambda) \\
&\leq -\log\left\{\left[\min_{\|w\|\leq\overline{w},\|v\|\leq\overline{v}} \exp(w'v)\right]\int_{\|v\|\leq\overline{v}} \frac{1}{1+\sum_{k=1}^{J_t}\exp(\check{\delta}_{kt}(\tilde{\pi}_t;\lambda)+w_{kt}'v)\tilde{\pi}_{kt}} dF(v;\lambda)\right\} \\
&\leq -\left[\min_{\|w\|\leq\overline{w},\|v\|\leq\overline{v}} (w'v)\right] - \log(\underline{\varepsilon}_0/4).
\end{aligned}$$

Thus, $\sup_{t,\lambda,\tilde{\pi}_t} \check{\delta}_{1t}(\tilde{\pi}_t;\lambda) < \infty$.

Now we show that $b_{\ell t}(\tilde{\pi}_t;\lambda)$ is uniformly bounded. By Cauchy-Schwarz inequality, it suffices to consider the ℓth element of $b_{\ell t}(\tilde{\pi}_t;\lambda)$:

$$\begin{aligned}
\tilde{\pi}_{\ell t}^{-2}\int \tilde{\pi}_{\ell t}(v)^2 dF(v;\lambda) &= \int \left(\frac{\exp(w_{\ell t}'v+\check{\delta}_{\ell t}(\tilde{\pi}_t;\lambda))}{1+\sum_{k=1}^{J_t}\exp(w_{kt}'v+\check{\delta}_{kt}(\tilde{\pi}_t;\lambda))\pi_{kt}}\right)^2 dF(v;\lambda) \\
&\leq \exp(2\check{\delta}_{\ell t}(\tilde{\pi}_t;\lambda))\int \exp(2w_{\ell t}'v)dF(v;\lambda).
\end{aligned}$$

Then by condition (ii) and $\sup_{t,\lambda,\tilde{\pi}_t} \check{\delta}_{\ell t}(\tilde{\pi}_t;\lambda) < \infty$, we have

$$\sup_t \sup_\lambda \sup_{\|\tilde{\pi}_t-\pi_t\|\leq e_2} \|\tilde{\pi}_{\ell t}^{-2}\int \tilde{\pi}_{\ell t}(v)^2 dF(v;\lambda)\| < \infty.$$

This shows the uniform boundedness of the elements of $b_{\ell t}(\tilde{\pi}_t;\lambda)$. □

E Approximate Log Share

In this section, we show some theoretical derivation that provides further support for the finiteness of $\bar{\iota}_\ell$ and the approximate value of $\underline{\iota}_u$. Lemma 10 shows that $\iota^*(n, n\pi)$ approaches

$(1-\pi)/2$ when $n\pi$ is large. Lemma 11 shows that $\iota^*(n, n\pi)$ approaches $n\pi$ when $n\pi$ is small. Both align well with the numerical results shown in Figure 2 and Table 2. Thus, we are confident that the conclusions regarding the approximate values of $\bar{\iota}_\ell$ and $\underline{\iota}_u$ drawn in Section 3.3 are correct, even though a fully rigorous theoretical proof is out of reach due to the lack of an analytical solution for the expectation of the logarithm of mean-shifted binomial or Poisson random variables. For two sequences of positive numbers a_n and b_n, we denote $a_n \propto b_n$ if $a_n/b_n = O(1)$ and $b_n/a_n = O(1)$.

Lemma 10. *Let q follow a binomial distribution with parameters (n, π). Consider a sequence of binomial distributions such that $\pi \propto n^{-\nu}$ with $\nu \in [0, 1)$. Then along this sequence we have:*

(a) *For any fixed constant $\iota > 0$,*

$$E[\log(q+\iota) - \log(n\pi)] = \frac{\iota}{n\pi} - \frac{1-\pi}{2n\pi} + o((n\pi)^{-1}).$$

where the $o((n\pi)^{-1})$ is uniform over ι in any bounded closed subinterval of $(0, \infty)$.

(b) $\iota^*(n, n\pi) - (1-\pi)/2 \to 0$.

Lemma 11. *Let q follow a binomial distribution with parameters (n, π). consider a sequence of binomial distributions such that $\pi \propto n^{-\nu}$ with $\nu > 1$. Then along this sequence we have*

$$\frac{\iota^*(n, n\pi)}{n\pi} \to 1.$$

Proof of Lemma 10. (a) First note that by the Chernoff's inequality, we have for any $c > 0$

$$\Pr(|q - n\pi| > (n\pi)^{0.5+c}) \le 2\exp\left(-\frac{(n\pi)^{2c}}{3}\right). \tag{122}$$

Decompose $E[\log(q+\iota) - \log(n\pi)]$ as

$$\begin{aligned} E[\log(q+\iota) - \log(n\pi)] &= E[(\log(q+\iota) - \log(n\pi))1\{|q-n\pi| \le (n\pi)^{0.5+c}\}] \\ &\quad + E[(\log(q+\iota) - \log(n\pi))||q-n\pi| > (n\pi)^{0.5+c}]\Pr(|q-n\pi| > (n\pi)^{0.5+c}) \end{aligned} \tag{123}$$

To bound each of the summands of the right-hand side of (123), first consider the derivation

$$\begin{aligned} \exp(-(n\pi)^{2c}/3) &= \exp(-(n\pi)^{2c}/3)n^2n^{-2} \\ &= \exp(-(n\pi)^{2c}/3 + 2\log n)n^{-2} = o(n^{-2}) \end{aligned} \tag{124}$$

where the last equality holds because $(n\pi)^{2c}/(3\log n) \propto n^{2c(1-\nu)}/(3\log n) \to \infty$.

Now consider the second summand of the right-hand side of (123). By (122), it is bounded by

$$\begin{aligned}&2\max\{|\log(\iota)-\log(n\pi)|,|\log(n+\iota)-\log(n\pi)|\}\exp(-(n\pi)^{2c}/3)\\&\leq C\log(n)\exp(-(n\pi)^{2c}/3)\\&=o(n^{-2}\log n)=o((n\pi)^{-1}),\end{aligned}\tag{125}$$

where C is a universal constant, the inequality holds because ι is a fixed positive constant and $\log(n\pi) \leq \log n$, the first equality holds due to (124) and the second equality holds because $n\pi \propto n^{(1-\nu)} = o(n^2/\log(n))$. It is easy to see that the $o(\cdot)$'s are uniform in ι on any compact interval on $(0,\infty)$.

Write $(q+\iota)/(n\pi) = 1 + (q - n\pi + \iota)/(n\pi)$, and use a Taylor series expansion of the logarithm around 1, and we can write the first summand of the right-hand side of (123) as

$$\begin{aligned}&E[(\log(q+\iota)-\log(n\pi))1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}]\\&=(n\pi)^{-1}E[(q-n\pi+\iota)1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}]\\&-2^{-1}(n\pi)^{-2}E[(q-n\pi+\iota)^2 1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}]\\&+2(3!)^{-1}(n\pi)^{-3}E[(1+\tilde{x})^{-3}(q-n\pi+\iota)^3 1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}],\end{aligned}\tag{126}$$

where $\tilde{x}$ is a value on the interval $[0,(q-n\pi+\iota)/(n\pi)]$. Consider the derivation:

$$\begin{aligned}(n\pi)^{-3}E[(1+\tilde{x})^{-3}(q-n\pi+\iota)^3 1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}]&\leq C(n\pi)^{-3}(\iota^3+(n\pi)^{1.5+3c})\\&=o((n\pi)^{-1}),\end{aligned}\tag{127}$$

where C is a universal constant, and the equality holds when we pick a $c \in (0, 1/6)$. Also consider the derivation

$$\begin{aligned}&E[(q-n\pi+\iota)^2 1\{|q-n\pi|\leq(n\pi)^{0.5+c}\}]\\&=E[(q-n\pi+\iota)^2]-E[(q-n\pi+\iota)^2||q-n\pi|>(n\pi)^{0.5+c}]\Pr(|q-n\pi|>(n\pi)^{0.5+c})\\&=E[(q-n\pi+\iota)^2]-O(n^2\exp(-(n\pi)^{2c}/3))\\&=E[(q-n\pi+\iota)^2]+o(1)\\&=n\pi(1-\pi)+\iota^2+o(1),\end{aligned}\tag{128}$$

where the second equality holds by (122) and $q \leq n$, the third equality holds by (124) and the last equality holds because q follows the binomial distribution with parameters (n,π).

By similar derivation, we have

$$E[(q - n\pi + \iota)1\{|q - n\pi| \leq (n\pi)^{0.5+c}\}] = \iota + o(1). \tag{129}$$

Combining (126)-(129), we have

$$E[(\log(q + \iota) - \log(n\pi))1\{|q - n\pi| \leq (n\pi)^{0.5+c}\}] = (n\pi)^{-1}(\iota - 2^{-1}(1 - \pi) + o(1)). \tag{130}$$

Equations (123), (125), and (130) together prove part (a) of the lemma.

(b) It is without loss of generality to assume that $\pi \to \pi_\infty$ for some $\pi_\infty \in [0, 1]$ as $n \to \infty$. (If not, we can consider subsubsequences of arbitrary subsequences of $\{n\}$ along with π converges. Such subsubsequences always exist because $[0, 1]$ is a compact set.) We consider two cases:

(i) $\pi_\infty = 1$. Suppose there exists a $\underline{c} > 0$ such that $\iota^*(n, n\pi) > \underline{c}$ infinitely often. Then by the monotonicity of the logarithm function, we have, infinitely often,

$$(n\pi)E[\log(q + \underline{c}) - \log(n\pi)] \leq (n\pi)E[\log(q + \iota^*(n, n\pi)) - \log(n\pi)] = 0, \tag{131}$$

where the equality holds by the definition of $\iota^*(n, n\pi)$. On the other hand, part(a) of the lemma implies that

$$(n\pi)[\log(q + \underline{c}) - \log(n\pi)] \to \underline{c} > 0. \tag{132}$$

This and (131) form a contradiction. Thus, there does not exist a $\underline{c} > 0$ such that $\iota^*(n, n\pi) > \underline{c}$ infinitely often. This implies that $\iota^*(n, n\pi) \to 0$, and in turn implies part (b) of the lemma.

(ii) $\pi_\infty \in (0, 1)$. Let $\underline{c} = (1 - \pi_\infty)/4$ and $\bar{c} = (1 - \pi_\infty)$. Suppose that $\iota^*(n, n\pi) < \underline{c}$ ($\iota^*(n, n\pi) > \bar{c}$) infinitely often. Then by the monotonicity of the logarithm function, we have, infinitely often, $(n\pi)E[\log(q+\underline{c}) - \log(n\pi) \geq 0$ ($(n\pi)E[\log(q+\bar{c}) - \log(n\pi) \leq 0$). But part(a) of the lemma implies that $(n\pi)E[\log(q+\underline{c}) - \log(n\pi)] \to \underline{c} - (1-\pi_\infty)/2 < 0$ ($(n\pi)E[\log(q+\bar{c}) - \log(n\pi)] \to \bar{c} - (1-\pi_\infty)/2 > 0$). These form a contradiction. Thus, $\iota^*(n, n\pi) \in [\underline{c}, \bar{c}]$ eventually. This, the compactness of the interval $[\underline{c}, \bar{c}]$, and part (a) of the lemma together imply that

$$(n\pi)[\log(q + \iota^*(n, n\pi)) - \log(n\pi)] - (\iota^*(n, n\pi) - (1 - \pi)/2) \to 0. \tag{133}$$

But $\log(q + \iota^*(n, n\pi)) - \log(n\pi) = 0$ by the definition of $\iota^*(n, n\pi)$. Thus,$\iota^*(n, n\pi) -$

$(1-\pi)/2 \to 0$, which shows part (b) of the lemma.

□

Proof of Lemma 11. By the definition of $\iota^*(n, n\pi)$, we have

$$\begin{aligned} 0 &= E[\log(q + \iota^*(n, n\pi)) - \log(n\pi)] \\ &= E[\log((q + \iota^*(n, n\pi))/(n\pi))] \\ &= (1-\pi)^n \log(\iota^*(n, n\pi)/(n\pi)) + (1-(1-\pi)^n) E[\log((q + \iota^*(n, n\pi))/(n\pi)) | q > 0]. \end{aligned} \tag{134}$$

Thus,

$$\log\left(\frac{\iota^*(n, n\pi)}{n\pi}\right) = -\frac{1-(1-\pi)^n}{(1-\pi)^n} E[\log((q + \iota^*(n, n\pi))/(n\pi)) | q > 0] \tag{135}$$

For $1 \le q \le n$, since $\iota^*(n, n\pi) \le n$, we have $0 < \log((q + \iota^*(n, n\pi))/(n\pi)) \le \log(2) - \log(\pi)$. Thus,

$$|E[\log((q + \iota^*(n, n\pi))/(n\pi)) | q > 0]| \le (\log(2) - \log(\pi)) \le 2|\log(\pi)|. \tag{136}$$

where the second inequality holds for large enough n since $n\pi \to 0$. Also consider

$$\begin{aligned} 1-(1-\pi)^n &= \binom{n}{1}\pi - \binom{n}{2}\pi^2 + \cdots + (-1)^{n+1}\binom{n}{n}\pi^n \\ &\le n\pi + (n\pi)^2 + \cdots + (n\pi)^n = \frac{n\pi(1-(n\pi)^n)}{1-n\pi} \end{aligned} \tag{137}$$

Thus, for large enough n, we have

$$\begin{aligned} \left|\log\left(\frac{\iota^*(n, n\pi)}{n\pi}\right)\right| &\le 2|\log \pi| \frac{n\pi(1-(n\pi)^n)}{1-n\pi-n\pi(1-(n\pi)^n)} \\ &= 2n\pi|\log(\pi)|(1+o(1)) \\ &= o(1), \end{aligned} \tag{138}$$

where the inequality holds by (136) and (137), and the equalities hold by $\pi \propto n^{-v}$ with $v > 1$. This proves the lemma. □