Supplementary Material for "Sample Empirical Likelihood and the Design-based Oracle Variable Selection Theory"

Puying Zhao[1], David Haziza[2] and Changbao Wu[3]

[1] *Yunnan University,* [2] *University of Ottawa and* [3] *University of Waterloo*

Abstract: This supplementary material contains technical details and proofs of the major theoretical results presented in the main paper and additional simulation results on point estimation, hypothesis tests and variable selection for linear regression and quantile regression models.

Key words and phrases: Design-based variable selection theory, Empirical likelihood ratio test, General hypothesis test, Nondifferentiable estimating functions, Quantile regression analysis, Survey weighted estimating equations.

Let $f_{\scriptscriptstyle N} = n_{\scriptscriptstyle B}/N$. One of the technical details is to replace π_i^{-1} in the constraints by $\pi_i^{-1} f_{\scriptscriptstyle N}$. The two versions of constraints $\sum_{i\in\mathcal{S}} p_i\{\pi_i^{-1} g_i(\theta)\} = 0$ and $\sum_{i\in\mathcal{S}} p_i\{\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)\} = 0$ are equivalent, but the latter version facilitates the usual asymptotic orders under Condition 3(ii). Let

$$l(\theta, \lambda) = n_{\scriptscriptstyle B}^{-1} \sum_{i\in\mathcal{S}} \log\{1 + \lambda^{\mathrm{T}} \pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)\}\,.$$

The maximum sample empirical likelihood estimator of $\theta_{\scriptscriptstyle N}$, which is the minimum point of $l_n(\theta, \lambda)$ given in (2.3) of the main paper, is equivalently given by

$$\hat{\theta}_{\scriptscriptstyle SEL} = \arg\min_{\theta\in\Theta} \sup_{\lambda\in\hat{\Lambda}_n(\theta)} l(\theta,\lambda)\,,$$

where $\hat{\Lambda}_n(\theta) = \{\lambda \mid \lambda^{\mathrm{T}}\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta) > -1, i \in \mathcal{S}\}$ for the given θ.

Let $\hat{\lambda}_{\scriptscriptstyle SEL} = \arg\sup_{\lambda\in\hat{\Lambda}_n(\hat{\theta}_{SEL})} l(\hat{\theta}_{\scriptscriptstyle EL}, \lambda)$. Let “w.p.a.1” denote “with probability approaching 1” and “$\xrightarrow{p}$” denote “converge in probability” under the design-based asymptotic framework. We also use “c” to denote a generic constant whenever the actual value of “c” is not a crucial part of the argument.

S1 Regularity Conditions

Let I_t denote the $t \times t$ identity matrix. Let $\|A\| = \{\mathrm{trace}(A^{\mathrm{T}}A)\}^{1/2}$ for any matrix or vector A. The following regularity conditions are used for the main theoretical results presented in the paper.

C1. The finite population parameter $\theta_{\scriptscriptstyle N} \in \Theta$ is the unique solution to $U_{\scriptscriptstyle N}(\theta) = 0$ and Θ is a compact set in the p-dimensional Euclidean space.

C2. There exists a function $U(\theta)$ such that $U_{\scriptscriptstyle N}(\theta) \to U(\theta)$ as $N \to \infty$, uniformly for all $\theta \in \Theta$. The limiting function $U(\theta)$ also satisfies the following conditions:

(i) There exists a unique solution to $U(\theta) = 0$, which is an interior point of Θ;

(ii) Uniformly for all $\theta \in \Theta$, the limiting function $U(\theta)$ is absolute continuous with first derivative $\Gamma(\theta) = \partial U(\theta)/\partial\theta$, and $\Gamma(\theta)$ has full column rank p for any $\theta \in \Theta$.

C3. The sampling design along with the expected sample size n_B satisfies

(i) $n_B = O(N^{\varrho})$ for some ϱ such that $1/2 < \varrho \leq 1$;

(ii) $c_1 < \pi_i N n_B^{-1} < c_2$, $i \in \mathcal{S}$ for some positive constants c_1 and c_2.

C4. The functions $U_N(\theta)$ and $U(\theta)$ satisfy

(i) For any sequence of positive numbers $\{\delta_N\}$ with $\delta_N = o_p(1)$,

$$\sup_{\theta\in\Theta(\delta_N)} \|[U_N(\theta) - U_N(\theta_N)] - [U(\theta) - U(\theta_N)]\| = o(N^{-1/2});$$

where $\Theta(\delta_N) = \big\{\theta \in \Theta : \|\theta - \theta_N\| \leq \delta_N\big\}$;

(ii) For any sequence $c_N = O(N^{-\eta})$ with $\eta \in (1/4, 1/2]$,

$$\sup_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^{N} \|g_i(\theta) - g_i(\theta + c_N)\| = O(|c_N|)\,;$$

(iii) There exists a positive constant c such that

$$\sup_{\theta\in\Theta(\delta)} \mathrm{Var}\Big\{[\hat{U}_N(\theta) - \hat{U}_N(\theta_N)] \mid \mathcal{F}_N\Big\} \leq c n_B^{-1}|\delta|\,,$$

for any $\delta > 0$, where $\Theta(\delta) = \big\{\theta \in \Theta : \ \|\theta - \theta_{N}\| \le \delta\big\}$.

C5. The finite population values $\mathcal{F}_{N}$, the estimating functions $g_i(\theta) = g(X_i, Y_i, \theta)$ and the sampling design satisfy

(i) $\max_{i\in\mathcal{S}} \sup_{\theta\in\Theta} \|g_i(\theta)\| = o_p(n_{B}^{1/2})$;

(ii) The Horvitz-Thompson estimator $\sum_{i\in\mathcal{S}} \pi_i^{-1} g_i(\theta_{N})$ is asymptotically normally distributed with mean zero and the design-based variance-covariance matrix at the order $O(n_{B}^{-1}N^2)$.

C6. For any vector Z satisfying $(1/N)\sum_{i=1}^{N}\|Z_i\|^{2+\sigma} < \infty$ with some $\sigma > 0$, $\mathrm{Var}(\hat{\mu}_Z \mid \mathcal{F}_{N}) \le c_0 n_{B}^{-1}(N-1)^{-1}\sum_{i=1}^{N}(Z_i - \mu_Z)(Z_i - \mu_Z)^{\mathrm{T}}$ for some constant c_0, where $\mu_Z = (1/N)\sum_{i=1}^{N} Z_i$ and $\hat{\mu}_Z = (1/N)\sum_{i\in\mathcal{S}} \pi_i^{-1} Z_i$.

Remark 1. Condition C1 ensures the identifiability of the parameter θ_{N}. Condition C2 specifies a smooth limiting function for the finite population function $U_{N}(\theta)$. We allow that $U_{N}(\theta)$ could be non-differentiable with respect to θ but impose a continuity assumption on its limiting function $U(\theta)$, while the existing survey sampling literatures (e.g., Chen and Kim (2014); Oguz-Alper and Berger (2016)) impose a continuity assumption directly on $g(X, Y, \theta)$ with respect to θ. Assume that the finite population $\{(X_i, Y_i), i = 1, \cdots, N\}$ is an independent and identically distributed sample from a superpopulation model with cumulative distribution function $F(x, y)$. Then in the model-based context, we have that $U(\theta) = E_{F}\{g(X, Y, \theta)\}$, which is differentiable with respect to θ regardless of whether the estimating function g is differentiable or non-differentiable. Here $E_{F}\{\cdot\}$ represents the expectation taken with respect to $F(x, y)$. Condition C3 is satisfied by most commonly used sampling designs. Conditions C4(i) and C4(ii) put a bound on the variation of population quantities, which is a typical condition in dealing with non-smooth estimating functions similar to those used in Bahadur representations and could be easily verified under a superpopulation model. Condition C4(iii) is about the correlation between two Horvitz-Thompson estimators at two close points of θ, which is similar to Condition 6 in Francisco and Fuller (1991) but without assuming a specific sampling design, and is a trivial condition when $g_i(\theta)$ is a smooth function of θ. It is used for the development of an approximation to the difference $\hat{U}_{N}(\theta) - \hat{U}_{N}(\theta_{N})$ on $\big\{\theta \in \Theta : \ \|\theta - \theta_{N}\| \leq \delta\big\}$ for non-differential estimating functions. Conditions (i), (ii) and (iii) in C5 are the regularity conditions commonly used for estimating equations with complex surveys. Condition C6 specifies that the variance-covariance matrix of the Horvitz-Thompson estimator under the given sampling design does not differ in terms of order of magnitude from the one under simple random sampling.

The following two conditions (Fan and Li (2001)) are assumed for the penalty function and the tuning parameter τ_n on design-based variable selection.

C7. As $n_{\scriptscriptstyle B} \to \infty$, $n_{\scriptscriptstyle B}^{1/2}\tau_n \to \infty$ and $\liminf_{n_B\to\infty} \liminf_{\theta\to 0^+} \tau_n^{-1} p'_{\tau_n}(\theta) > 0$.

C8. $\max_{j\in\mathcal{A}} p'_\tau(|\theta_{N[j]}|) = o(n_{\scriptscriptstyle B}^{-1/2})$ and $\max_{j\in\mathcal{A}} p''_\tau(|\theta_{N[j]}|) = o(1)$.

S2 Lemmas

Lemma 1. *Suppose that Conditions C1, C3 and C5 hold. Then*

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n,i\in\mathcal{S}} |\lambda^{\mathrm{T}}\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)| = o_p(1)\,,$$

where $\Lambda_n = \{\lambda \mid \|\lambda\| \le c n_{\scriptscriptstyle B}^{-1/2}\}$ *for a given* $c > 0$*. In addition, w.p.a.1,* $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ *for all* $\theta \in \Theta$*.*

Proof. It can be shown that $\max_{i\in\mathcal{S}} \sup_{\theta\in\Theta} \|\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)\| = o_p(n_{\scriptscriptstyle B}^{1/2})$ by Conditions C3(ii) and C5(i). The use of Cauchy-Schwarz inequality leads to

$$\sup_{\theta\in\Theta,\lambda\in\Lambda_n,i\in\mathcal{S}} |\lambda^{\mathrm{T}}\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)| \le \|\lambda\| \max_{i\in\mathcal{S}} \sup_{\theta\in\Theta} \|\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta)\| = o_p(1)\,.$$

Moreover, w.p.a.1, $\lambda^{\mathrm{T}}\pi_i^{-1} f_{\scriptscriptstyle N} g_i(\theta) \in (-1,\infty)$ for all $\theta \in \Theta$ and $\|\lambda\| \le n_{\scriptscriptstyle B}^{-1/2}$.

□

Lemma 2. *Suppose that Conditions C1, C3 and C5 hold,* $\bar{\theta} \in \Theta$*,* $\bar{\theta} \xrightarrow{p} \theta_{\scriptscriptstyle N}$ *and* $\|\hat{U}_{\scriptscriptstyle N}(\bar{\theta})\| = O_p(n^{-1/2})$*. Then,* $\bar{\lambda} = \arg\sup_{\lambda\in\hat{\Lambda}_n(\bar{\theta})} l(\bar{\theta},\lambda)$ *exists w.p.a.1,* $\bar{\lambda} = O_p(n_{\scriptscriptstyle B}^{-1/2})$*, and* $\sup_{\lambda\in\hat{\Lambda}_n(\bar{\theta})} l(\bar{\theta},\lambda) \le O_p(n_{\scriptscriptstyle B}^{-1})$*.*

Proof. Let $\mathcal{D}_n(\theta) = n_{_B}^{-1}\sum_{i\in\mathcal{S}}\pi_i^{-2}f_{_N}^2 g_i(\theta)g_i(\theta)^{\rm T}$. Applying a second order Taylor series expansion, we have $l(\bar\theta,\lambda) = \lambda^{\rm T}\hat U_{_N}(\bar\theta) - \frac{1}{2}\lambda^{\rm T}\mathcal{D}_n(\bar\theta)\lambda$. This further leads to the first order condition: $\hat U_{_N}(\bar\theta) - \mathcal{D}_n(\bar\theta)\lambda = 0$. By Condition C5(iii), $\|\mathcal{D}_n(\bar\theta) - W\| = o_p(1)$, where $W = n_{_B}N^{-2}\sum_{i=1}^N \pi_i^{-1}g_i(\theta_{_N})g_i(\theta_{_N})^{\rm T}$. It can be seen that, w.p.a.1, the smallest eigenvalue of $\mathcal{D}_n(\bar\theta)$ is bounded away from zero due to the nonsingularity of W. We conclude that, w.p.a.1, $\bar\lambda = \arg\sup_{\lambda\in\hat\Lambda_n(\bar\theta)} l(\bar\theta,\lambda)$ exists.

By Lemma 1, we further have $|\dot\lambda^{\rm T}\pi_i^{-1}f_{_N}g_i(\bar\theta)| = o_p(1)$, uniformly over $i\in\mathcal{S}$. It follows that $\max_{i\in\mathcal{S}}\{1+\dot\lambda^{\rm T}\pi_i^{-1}f_{_N}g_i(\bar\theta)\}^{-2} > 1/2$ w.p.a.1. Applying a second order Taylor series expansion with respect to λ, we have that for some $\dot\lambda$ on the line segment between $\bar\lambda$ and 0,

$$\begin{aligned} l(\bar\theta,\bar\lambda) &= \bar\lambda^{\rm T}\hat U_{_N}(\bar\theta) - \frac{1}{2}\bar\lambda^{\rm T}\left(\frac{1}{n_{_B}}\sum_{i\in\mathcal{S}}^{n}\frac{\pi_i^{-2}f_{_N}^2 g_i(\bar\theta)g_i(\bar\theta)^{\rm T}}{\{1+\dot\lambda^{\rm T}\pi_i^{-1}f_{_N}g_i(\bar\theta)\}^2}\right)\bar\lambda \\ &\le \|\bar\lambda^{\rm T}\|\|\hat U_{_N}(\bar\theta)\| - c\|\bar\lambda^{\rm T}\|^2. \end{aligned}$$

Since $\bar\lambda$ is the maximizer, $l(\bar\theta,\bar\lambda) \ge l(\bar\theta,0) = 0$. This, coupled with the assumption that $\|\hat U_{_N}(\bar\theta)\| = O_p(n^{-1/2})$, implies that $\|\bar\lambda\| = O_p(n_{_B}^{-1/2})$. We also conclude that $l(\bar\theta,\bar\lambda) = \sup_{\lambda\in\hat\Lambda_n(\bar\theta)} l(\bar\theta,\lambda) \le O_p(n_{_B}^{-1})$. □

Lemma 3. *Suppose that Conditions C1, C3 and C5 hold. Then* $\|\hat U_{_N}(\hat\theta_{_{SEL}})\| = O_p(n_{_B}^{-1/2})$.

Proof. Let $\tilde{\lambda} = n_{B}^{-1/2}\hat{U}_{N}(\hat{\theta}_{SEL})/\|\hat{U}_{N}(\hat{\theta}_{SEL})\|$. It follows from Lemma 1 that

$$\max_{i\in\mathcal{S}} |\tilde{\lambda}^{\mathrm{T}}\pi_i^{-1}f_{N}g_i(\hat{\theta}_{SEL})| \ \xrightarrow{p}\ 0$$

and $\tilde{\lambda} \in \hat{\Lambda}_n(\hat{\theta}_{SEL})$. Moreover, by the Cauchy–Schwarz inequality and Condition C5(iii), it can be shown that

$$n_{B}^{-1}\sum_{i\in\mathcal{S}}\pi_i^{-2}f_{N}^2 g_i(\hat{\theta}_{SEL})g_i(\hat{\theta}_{SEL})^{\mathrm{T}} \le n_{B}^{-1}\sum_{i\in\mathcal{S}}\pi_i^{-2}f_{N}^2\sup_{\theta\in\Theta}\|g_i(\theta)\|^2 I_r \xrightarrow{p} cI_r\,.$$

By using the Taylor series expansion, we have

$$\begin{aligned} l(\hat{\theta}_{SEL},\tilde{\lambda}) &= \tilde{\lambda}^{\mathrm{T}}\hat{U}_{N}(\hat{\theta}_{SEL}) - \frac{1}{2}\tilde{\lambda}^{\mathrm{T}}\left(\frac{1}{n_B}\sum_{i\in\mathcal{S}}^{n}\frac{\pi_i^{-2}f_{N}^2 g_i(\hat{\theta}_{SEL})g_i(\hat{\theta}_{SEL})^{\mathrm{T}}}{\{1+\dot{\lambda}^{\mathrm{T}}\pi_i^{-1}f_{N}g_i(\hat{\theta}_{SEL})\}^2}\right)\tilde{\lambda} \\ &\ge\ n_{B}^{-1/2}\|\hat{U}_{N}(\hat{\theta}_{SEL})\| - (c/4)n_{B}^{-1}. \end{aligned}$$

This, together with the fact that $\hat{\theta}_{SEL}$ and $\hat{\lambda}_{SEL}$ are a saddle point, implies that

$$\begin{aligned} n_{B}^{-1/2}\|\hat{U}_{N}(\hat{\theta}_{SEL})\| - (c/4)n_{B}^{-1} &\le l(\hat{\theta}_{SEL},\tilde{\lambda}) \le l(\hat{\theta}_{SEL},\hat{\lambda}_{SEL}) \\ &\le \sup_{\lambda\in\hat{\Lambda}_n(\theta_N)} l(\theta_{N},\lambda) \\ &\le O_p(n_{B}^{-1}). \end{aligned}$$

The two sets of inequalities lead to $\|\hat{U}_{N}(\hat{\theta}_{SEL})\| \le O_p(n_{B}^{-1/2})$.

Now, we consider $\bar{\lambda} = \varepsilon_n\hat{U}_{N}(\hat{\theta}_{SEL})$ with any $\varepsilon_n \to 0$. Similarly, we have that

$$\varepsilon_n\|\hat{U}_{N}(\hat{\theta}_{SEL})\|^2 - (c/4)\varepsilon_n^2\|\hat{U}_{N}(\hat{\theta}_{SEL})\|^2 \le O_p(n_{B}^{-1}).$$

It can be shown that $\varepsilon_n \|\hat{U}_{\scriptscriptstyle N}(\hat{\theta}_{\scriptscriptstyle SEL})\|^2 = O_p(n_{\scriptscriptstyle B}^{-1})$ by noting that $1-(c/4)\varepsilon_n > 0$ for $n_{\scriptscriptstyle B}$ large enough. Then we can show $\|\hat{U}_{\scriptscriptstyle N}(\hat{\theta}_{\scriptscriptstyle SEL})\| = O_p(n_{\scriptscriptstyle B}^{-1/2})$. □

Lemma 4. *Suppose that Conditions C1–C6 hold. Then, for any sequence of positive numbers* $\{\delta_{\scriptscriptstyle N}\}$ *with* $\delta_{\scriptscriptstyle N} = o_p(1)$*,*

$$\sup_{\theta\in\Theta(\delta_N)} \|[\hat{U}_{\scriptscriptstyle N}(\theta) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})] - [U(\theta) - U(\theta_{\scriptscriptstyle N})]\| = o_p(n_{\scriptscriptstyle B}^{-1/2})\,,$$

where $\Theta(\delta_{\scriptscriptstyle N}) = \Big\{\theta \in \Theta : \ \|\theta - \theta_{\scriptscriptstyle N}\| \le \delta_{\scriptscriptstyle N}\Big\}$.

Proof. Note that $\hat{U}_{\scriptscriptstyle N}(\theta) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) - U(\theta) + U(\theta_{\scriptscriptstyle N}) = A_{\scriptscriptstyle N}(\theta) + B_{\scriptscriptstyle N}(\theta)$, where $A_{\scriptscriptstyle N}(\theta) = \hat{U}_{\scriptscriptstyle N}(\theta) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) - U_{\scriptscriptstyle N}(\theta) + U_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})$ and $B_{\scriptscriptstyle N}(\theta) = U_{\scriptscriptstyle N}(\theta) - U_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) - U(\theta) + U(\theta_{\scriptscriptstyle N})$. By Conditions C3 and C4(i), it can be shown that

$$\sup_{\theta\in\Theta(\delta_N)} \|B_{\scriptscriptstyle N}(\theta)\| = o(n_{\scriptscriptstyle B}^{-1/2}).$$

We now consider the asymptotic property of $\|A_{\scriptscriptstyle N}(\theta)\|$. We have

$$\begin{aligned} &\mathrm{E}[\|A_{\scriptscriptstyle N}(\theta)\|^2 \mid \mathcal{F}_{\scriptscriptstyle N}] = \mathrm{trace}\Big\{\mathrm{E}[A_{\scriptscriptstyle N}(\theta)A_{\scriptscriptstyle N}(\theta)^{\mathrm{T}} \mid \mathcal{F}_{\scriptscriptstyle N}]\Big\} \\ &= \mathrm{E}[A_{\scriptscriptstyle N}(\theta)^{\mathrm{T}} \mid \mathcal{F}_{\scriptscriptstyle N}]\mathrm{E}[A_{\scriptscriptstyle N}(\theta) \mid \mathcal{F}_{\scriptscriptstyle N}] + \mathrm{trace}\Big\{\mathrm{Var}[A_{\scriptscriptstyle N}(\theta) \mid \mathcal{F}_{\scriptscriptstyle N}]\Big\}. \end{aligned}$$

It can be seen that $\mathrm{E}\{A_{\scriptscriptstyle N}(\theta) \mid \mathcal{F}_{\scriptscriptstyle N}\} = 0$ and $\mathrm{Var}\{A_{\scriptscriptstyle N}(\theta) \mid \mathcal{F}_{\scriptscriptstyle N}\} = \mathrm{Var}\{[\hat{U}_{\scriptscriptstyle N}(\theta) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})] \mid \mathcal{F}_{\scriptscriptstyle N}\}$ for all θ in $\Theta(\delta_{\scriptscriptstyle N})$. In addition, it follows from Condition C4(iii) that

$$\mathrm{Var}\Big\{[\hat{U}_{\scriptscriptstyle N}(\theta) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})] \mid \mathcal{F}_{\scriptscriptstyle N}\Big\} \le c n_{\scriptscriptstyle B}^{-1} o(1),$$

uniformly for θ in $\Theta(\delta_{\scriptscriptstyle N})$ with $\delta_{\scriptscriptstyle N} = o(1)$. Then $\mathrm{E}[\|A_{\scriptscriptstyle N}(\theta)\|^2 \mid \mathcal{F}_{\scriptscriptstyle N}] = o_p(n_{\scriptscriptstyle B}^{-1})$

uniformly for θ in $\Theta(\delta_{\scriptscriptstyle N})$. This leads to $\sup_{\theta\in\Theta(\delta_N)}\|A_{\scriptscriptstyle N}(\theta)\| = o_p(n_{\scriptscriptstyle B}^{-1/2})$, which completes the proof. □

Lemma 5. *Suppose that Conditions C1–C6 hold. Then*

$$\sup_{\theta\in\Theta}\|\hat U_{\scriptscriptstyle N}(\theta)-U_{\scriptscriptstyle N}(\theta)\| = o_p(1)\,.$$

Proof. We prove the lemma by using the covering approach (van der Vaart and Wellner (1996); Wang and Opsomer (2011)). Let $1/2<\delta<\varrho$. Since Θ is compact, we can partition Θ into N^δ subsets such that $\Theta=\cup_{j=1}^{N^\delta}\Theta_j$ and $\|\theta_j'-\theta_j\|\le c_{\scriptscriptstyle N}$ for any $\theta_j',\theta_j\in\Theta_j$, where $c_{\scriptscriptstyle N}=O(N^{-\delta})$. Then, for any $\theta_j\in\Theta_j, j=1,2,\cdots,N^\delta$, we have

$$\begin{aligned}
&\sup\nolimits_{\theta\in\Theta}\|\hat U_{\scriptscriptstyle N}(\theta)-U_{\scriptscriptstyle N}(\theta)\|\\
\le\;& \max\nolimits_j\|\hat U_{\scriptscriptstyle N}(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\|\\
&+\max\nolimits_j\sup\nolimits_{\theta\in\Theta_j}\|\{\hat U_{\scriptscriptstyle N}(\theta)-U_{\scriptscriptstyle N}(\theta)\}-\{\hat U_{\scriptscriptstyle N}(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\}\|\,.
\end{aligned}$$

By Condition C6, we have

$$\begin{aligned}
&\mathrm{Var}\{\hat U_{\scriptscriptstyle N}(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\mid\mathcal F_{\scriptscriptstyle N}\}\\
\le\;& c_0\frac{1}{n_{\scriptscriptstyle B}}\frac{1}{N-1}\sum_{i=1}^N\{g_i(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\}\{g_i(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\}^{\mathrm T}\\
=\;& O(n_{\scriptscriptstyle B}^{-1})\,.
\end{aligned}$$

Since $\delta<\varrho$, we have that for any $\epsilon>0$,

$$\begin{aligned}
\Pr\Big(\max\nolimits_j\|\hat U_{\scriptscriptstyle N}(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\|\ge\epsilon\mid\mathcal F_{\scriptscriptstyle N}\Big) &\le \sum_{j=1}^{N^\delta}\frac{E\{\|\hat U_{\scriptscriptstyle N}(\theta_j)-U_{\scriptscriptstyle N}(\theta_j)\|^2\}}{\epsilon^2}\\
&= O(N^{\delta-\varrho})\,,
\end{aligned}$$

with $O(N^{\delta-\varrho}) \to 0$. Let $I_i = I(i \in \mathcal{S})$ be the sample inclusion indicators. By Condition C3(ii), we have $|I_i/\pi_i - 1| \leq cN/n_B$ for some $c > 0$ uniformly over i. By Conditions C3(i) and C4(ii), we have that

$$\begin{aligned}
& \max_j \sup_{\theta \in \Theta_j} \|\{\hat{U}_N(\theta) - U_N(\theta)\} - \{\hat{U}_N(\theta_j) - U_N(\theta_j)\}\| \\
= \ & \max_j \sup_{\theta \in \Theta_j} \left\| \frac{1}{N} \sum_{i=1}^{N} \left(\frac{I_i}{\pi_i} - 1 \right) \{g_i(\theta) - g_i(\theta_j)\}] \right\| \\
\leq \ & c \frac{N}{n_B} \max_j \sup_{\theta \in \Theta_j} \frac{1}{N} \sum_{i=1}^{N} \|\{g_i(\theta) - g_i(\theta_j)\}]\| \\
= \ & O(N^{1-\varrho-\delta}),
\end{aligned}$$

and $O(N^{1-\varrho-\delta}) \to 0$. Combining all the above arguments, we conclude that $\sup_{\theta \in \Theta} \|\hat{U}_N(\theta) \ -U_N(\theta)\| = o_p(1)$. □

S3 Proofs of Theorems

Proof of Theorem 1. We first consider the consistency of the maximum sample empirical likelihood estimator $\hat{\theta}_{SEL}$. Note that $\{\theta : \|\theta - \theta_N\| \geq \epsilon\} = \Theta - \{\theta : \|\theta - \theta_N\| < \epsilon\}$ is also a compact subset of Θ for any $\epsilon > 0$. Thus, there exists $\theta_1 \in \{\theta : \|\theta - \theta_N\| \geq \epsilon\}$ such that

$$\inf_{\theta : \|\theta - \theta_N\| \geq \epsilon} \|U_N(\theta)\| = \|U_N(\theta_1)\|.$$

Since θ_N is the unique solution to $U_N(\theta) = 0$ and $\theta_1 \neq \theta_N$, $\|U_N(\theta_1)\| > 0$. Thus $\inf_{\theta : \|\theta - \theta_N\| \geq \epsilon} \|U_N(\theta)\| > 0$ for all $\epsilon > 0$. This implies that for every

$\epsilon > 0$ there exists a number $\eta(\epsilon) > 0$ such that $\|U_N(\theta)\| \geq \eta(\epsilon) > 0$ for every θ with $\|\theta - \theta_N\| > \epsilon$. Thus, the event $\{\|\theta - \theta_N\| > \epsilon\}$ is contained in the event $\{\|U_N(\theta)\| \geq \eta(\epsilon) > 0\}$, and $\Pr(\|\hat{\theta}_{SEL} - \theta_N\| > \epsilon \mid \mathcal{F}_N) \leq \Pr(\|U_N(\hat{\theta}_{SEL})\| \geq \eta(\epsilon) \mid \mathcal{F}_N)$ for all $\epsilon > 0$. It suffices to show that $\|U_N(\hat{\theta}_{SEL})\| = o_p(1)$. Applying the triangle inequality and by Lemma 3 and Lemma 5, it can be show that

$$\|U_N(\hat{\theta}_{EL})\| \leq \|U_N(\hat{\theta}_{EL}) - \hat{U}_N(\hat{\theta}_{EL})\| + \|\hat{U}_N(\hat{\theta}_{EL})\| = o_p(1).$$

The consistency of $\hat{\theta}_{SEL}$ then follows.

The proof of the asymptotic normality of $\hat{\theta}_{SEL}$ can be carried out in three steps.

Step 1. Show that $\|\hat{\theta}_{SEL} - \theta_N\| = O_p(n_B^{-1/2})$. Using triangle inequalities, we have that

$$\|U(\hat{\theta}_{SEL}) - U(\theta_N)\| \leq \|U(\hat{\theta}_{SEL}) - U(\theta_N) - \hat{U}_N(\hat{\theta}_{SEL}) + \hat{U}_N(\theta_N)\| \\ + \|\hat{U}_N(\hat{\theta}_{SEL})\| + \|\hat{U}_N(\theta_N)\|.$$

It follows from Lemma 3 that $\|\hat{U}_N(\hat{\theta}_{SEL})\| = O_p(n_B^{-1/2})$ and, by Conditions C2(ii) and C5(iii), $\|U(\theta_N)\| = O_p(n_B^{-1/2})$ and $\|\hat{U}_N(\theta_N)\| = O_p(n_B^{-1/2})$. Lemma 4 implies that

$$\|\hat{U}_N(\hat{\theta}_{SEL}) - \hat{U}_N(\theta_N) - U(\hat{\theta}_{SEL}) + U(\theta_N)\| \leq (1 + n_B^{1/2}\|\hat{\theta}_{SEL} - \theta_N\|)o_p(n_B^{-1/2}).$$

Hence,

$$\|U(\hat{\theta}_{SEL}) - U(\theta_N)\| \leq (1 + n_B^{1/2}\|\hat{\theta}_{SEL} - \theta_N\|)o_p(n_B^{-1/2}) + O_p(n_B^{-1/2})\,.$$

It follows that $\|U(\hat{\theta}_{SEL}) - U(\theta_N)\| \geq C\|\hat{\theta}_{SEL} - \theta_N\|$ because $U(\theta)$ is differentiable at θ_N. We have

$$n_B^{1/2}\|\hat{\theta}_{SEL} - \theta_N\| \leq (1 + n_B^{1/2}\|\hat{\theta}_{SEL} - \theta_N\|)o_p(1) + O_p(1)\,,$$

which leads to $(1 - o_p(1))n_B^{1/2}\|\hat{\theta}_{SEL} - \theta_N\| \leq O_p(1)$ and $\|\hat{\theta}_{SEL} - \theta_N\| = O_p(n_B^{-1/2})$.

Step 2. Show that $l(\theta, \lambda)$ can be approximated by the quadratic function

$$L(\theta, \lambda) = [\Gamma(\theta_N)(\theta - \theta_N)]^{\mathrm{T}}\lambda + \hat{U}_N(\theta_N)^{\mathrm{T}}\lambda - \frac{1}{2}\lambda^{\mathrm{T}}W\lambda$$

when (θ, λ) is in the neighbourhood of $(\theta_N, 0)$, where

$$W = n_B N^{-2}\sum_{i=1}^{N}\pi_i^{-1}g_i(\theta_N)g_i(\theta_N)^{\mathrm{T}}.$$

This can be achieved by showing that

$$|l(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL}) - L(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL})| = o_p(n_B^{-1})\,. \tag{S3.1}$$

The second order Taylor series expansion for $l(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL})$ at $\lambda = 0$ gives

$$\begin{aligned} l(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL}) &= \hat{\lambda}_{SEL}^{\mathrm{T}}\hat{U}_N(\hat{\theta}_{SEL}) \\ &\quad -\frac{1}{2}\hat{\lambda}_{SEL}^{\mathrm{T}}\left(\frac{1}{n_B}\sum_{i\in\mathcal{S}}^{n}\frac{\pi_i^{-2}f_N^2 g_i(\hat{\theta}_{SEL})g_i(\hat{\theta}_{SEL})^{\mathrm{T}}}{\{1 + \dot{\lambda}^{\mathrm{T}}\pi_i^{-1}f_N g_i(\hat{\theta}_{SEL})\}^2}\right)\hat{\lambda}_{SEL} \end{aligned}$$

for some $\dot{\lambda}$ on the line segment between $\hat{\lambda}_{SEL}$ and 0. This leads to

$$\begin{aligned}
& |l(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL}) - L(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL})| \\
\leq\; & |[\hat{U}_N(\hat{\theta}_{SEL}) - \hat{U}_N(\theta_N) - \Gamma(\theta_N)(\hat{\theta}_{SEL} - \theta_N)]^{\mathrm{T}} \hat{\lambda}_{SEL}| \\
& + \frac{1}{2}\left| \hat{\lambda}_{SEL}^{\mathrm{T}} \left(\frac{1}{n_B} \sum_{i \in \mathcal{S}} \frac{\pi_i^{-2} f_N^2 g_i(\hat{\theta}_{SEL}) g_i(\hat{\theta}_{SEL})^{\mathrm{T}}}{\{1 + \dot{\lambda}^{\mathrm{T}} \pi_i^{-1} f_N g_i(\hat{\theta}_{SEL})\}^2} - W \right) \hat{\lambda}_{SEL} \right| .
\end{aligned}$$

It can be shown that

$$\left\| \frac{1}{n_B} \sum_{i \in \mathcal{S}} \pi_i^{-2} f_N^2 g_i(\hat{\theta}_{SEL}) g_i(\hat{\theta}_{SEL})^{\mathrm{T}} - W \right\| = o_p(1) .$$

This, together with Lemmas 1 and 2, implies that

$$\left| \hat{\lambda}_{SEL}^{\mathrm{T}} \left(\frac{1}{n_B} \sum_{i \in \mathcal{S}} \frac{\pi_i^{-2} f_N^2 g_i(\hat{\theta}_{SEL}) g_i(\hat{\theta}_{SEL})^{\mathrm{T}}}{1 + \dot{\lambda}^{\mathrm{T}} \pi_i^{-1} f_N g_i(\hat{\theta}_{SEL})} - W \right) \hat{\lambda}_{SEL} \right| \leq \|\hat{\lambda}_{SEL}\|^2 o_p(1) = o_p(n_B^{-1}) .$$

We also have that

$$\begin{aligned}
& \|\hat{U}_N(\hat{\theta}_{SEL}) - \hat{U}_N(\theta_N) - \Gamma(\theta_N)(\hat{\theta}_{SEL} - \theta_N)\| \\
\leq\; & \|\hat{U}_N(\hat{\theta}_{SEL}) - \hat{U}_N(\theta_N) - U(\hat{\theta}_{SEL}) + U(\theta_N)\| \\
& + \|U(\hat{\theta}_{SEL}) - U(\theta_N) - \Gamma(\theta_N)(\hat{\theta}_{SEL} - \theta_N)\| \\
\leq\; & (1 + n_B^{1/2} \|\hat{\theta} - \theta_N\|) o_p(n_B^{-1/2}) + o_p(\|\hat{\theta}_{SEL} - \theta_N\|) \\
=\; & o_p(n_B^{-1/2}) ,
\end{aligned}$$

which further leads to

$$|[\hat{U}_N(\hat{\theta}_{SEL}) - \hat{U}_N(\theta_N) - \Gamma(\theta_N)(\hat{\theta}_{SEL} - \theta_N)]^{\mathrm{T}} \hat{\lambda}_{SEL}| = o_p(n_B^{-1}) .$$

The statement in equation (S3.1) follows immediately.

We now consider the alternative problem $\min_{\theta\in\Theta}\sup_{\lambda\in\mathcal{R}} L(\theta,\lambda)$. Since $L(\theta,\lambda)$ is concave in λ and Θ is compact, the first-order conditions (i.e., setting each of the derivatives to be zero) for an interior global maximum are satisfied at $(\tilde{\theta}^{\mathrm{T}},\tilde{\lambda}^{\mathrm{T}})^{\mathrm{T}}$ and are given by

$$\Gamma(\theta_{N})^{\mathrm{T}}\tilde{\lambda}=0\,,\qquad \Gamma(\theta_{N})(\tilde{\theta}-\theta_{N})+\hat{U}_{N}(\theta_{N})-W\tilde{\lambda}=0\,. \tag{S3.2}$$

The two systems of equations can be combined and rewritten as

$$\begin{pmatrix} 0 \\ -\hat{U}_{N}(\theta_{N}) \end{pmatrix} - \begin{pmatrix} 0 & \Gamma(\theta_{N})^{\mathrm{T}} \\ \Gamma(\theta_{N}) & -W \end{pmatrix} \begin{pmatrix} \tilde{\theta}-\theta_{N} \\ \tilde{\lambda}-0 \end{pmatrix} = 0\,.$$

Let $\Gamma=\Gamma(\theta_{N})$, $\Sigma=(\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1}$, $H=\Sigma\Gamma^{\mathrm{T}}W^{-1}$ and $P=W^{-1}-W^{-1}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1}$. Using the result on the inverse of a block matrix, we have

$$\begin{pmatrix} \tilde{\theta}-\theta_{N} \\ \tilde{\lambda}-0 \end{pmatrix} = \begin{pmatrix} \Sigma & H \\ H^{\mathrm{T}} & -P \end{pmatrix} \begin{pmatrix} 0 \\ -\hat{U}_{N}(\theta_{N}) \end{pmatrix}.$$

It follows that $\tilde{\theta}-\theta_{N}=-H\hat{U}_{N}(\theta_{N})=-(\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1}\Gamma^{\mathrm{T}}W^{-1}\hat{U}_{N}(\theta_{N})$. The asymptotic normality of $\tilde{\theta}$ follows from Condition C5(iii) and the asymptotic variance-covariance matrix of $\tilde{\theta}$ is given by

$$V_1=(\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1}\Gamma^{\mathrm{T}}W^{-1}\Omega W^{-1}\Gamma(\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1}\,,$$

where $\Omega=\mathrm{Var}\big\{\hat{U}_{N}(\theta_{N})\mid\mathcal{F}_{N}\big\}=\mathrm{Var}\big\{N^{-1}\sum_{i\in\mathcal{S}}\pi_i^{-1}g(X_i,Y_i,\theta_{N})\mid\mathcal{F}_{N}\big\}$.

Step 3. We show that $\hat{\theta}_{SEL}-\tilde{\theta}=o_p(n_{B}^{-1/2})$, i.e., $\hat{\theta}_{SEL}$ and $\tilde{\theta}$ are asymptotically equivalent. By the differentiability of the limiting function $U(\theta)$ at

$\theta_{\scriptscriptstyle N}$, it can be shown that

$$\|U(\tilde{\theta}) - U(\theta_{\scriptscriptstyle N})\| \leq \|\Gamma(\theta_{\scriptscriptstyle N})(\tilde{\theta} - \theta_{\scriptscriptstyle N})\| + o_p(\|\tilde{\theta} - \theta_{\scriptscriptstyle N}\|) = O_p(n^{-1/2})\,.$$

Using similar arguments to the proof of Lemma 4, together with Assumption 2(ii), we have

$$\begin{aligned} \|\hat{U}_{\scriptscriptstyle N}(\tilde{\theta})\| &\leq \|\hat{U}_{\scriptscriptstyle N}(\tilde{\theta}) - \hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) - U(\tilde{\theta}) + U(\theta_{\scriptscriptstyle N})\| \\ &\quad + \|\hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})\| + \|U(\tilde{\theta}) - U(\theta_{\scriptscriptstyle N})\| \\ &\leq (1 + n_{\scriptscriptstyle B}^{1/2}\|\tilde{\theta} - \theta_{\scriptscriptstyle N}\|)o_p(n_{\scriptscriptstyle B}^{-1/2}) + O_p(n_{\scriptscriptstyle B}^{-1/2}) \\ &= O_p(n_{\scriptscriptstyle B}^{-1/2})\,. \end{aligned}$$

It follows from similar arguments to (S3.1) that $|l(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL}) - L(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL})| = o_p(n_{\scriptscriptstyle B}^{-1})$. Noting that $l(\hat{\theta}_{\scriptscriptstyle SEL}, \hat{\lambda}_{\scriptscriptstyle SEL}) \leq l(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL})$, we have

$$L(\hat{\theta}_{\scriptscriptstyle SEL}, \hat{\lambda}_{\scriptscriptstyle SEL}) - o_p(n_{\scriptscriptstyle B}^{-1}) \leq l(\hat{\theta}_{\scriptscriptstyle SEL}, \hat{\lambda}_{\scriptscriptstyle SEL}) \leq l(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL}) \leq L(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL}) + o_p(n_{\scriptscriptstyle B}^{-1}).$$

Since $L(\hat{\theta}_{\scriptscriptstyle SEL}, \hat{\lambda}_{\scriptscriptstyle EL}) \geq L(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL})$, we conclude that $L(\hat{\theta}_{\scriptscriptstyle SEL}, \hat{\lambda}_{\scriptscriptstyle SEL}) - L(\tilde{\theta}, \hat{\lambda}_{\scriptscriptstyle SEL}) = o_p(n_{\scriptscriptstyle B}^{-1})$, which further leads to $[-\Gamma(\theta_{\scriptscriptstyle N})(\hat{\theta}_{\scriptscriptstyle SEL} - \tilde{\theta})]^{\mathrm{T}}\hat{\lambda}_{\scriptscriptstyle SEL} = o_p(n_{\scriptscriptstyle B}^{-1})$. It follows that $\hat{\theta}_{\scriptscriptstyle SEL} - \tilde{\theta} = o_p(n_{\scriptscriptstyle B}^{-1/2})$ since $\Gamma(\theta_{\scriptscriptstyle N})$ has full rank and $\hat{\lambda}_{\scriptscriptstyle SEL} = O_p(n_{\scriptscriptstyle B}^{-1/2})$. □

Proof of Corollary 1. Part (i) of Corollary 1 is straightforward since both Γ and W are invertible $p \times p$ matrices when $r = p$.

Under single-stage PPS sampling with replacement, the Horvitz-Thompson estimator $\hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) = N^{-1}\sum_{i\in\mathcal{S}}\pi_i^{-1}g_i(\theta_{\scriptscriptstyle N})$ can be re-written as the Hanson-

Hurwitz estimator $N^{-1}n^{-1}\sum_{i\in\mathcal{S}} z_i^{-1}g_i(\theta_{_N})$ with $\pi_i = nz_i$. Treating $\theta_{_N}$ as a known quantity, the design-unbiased variance estimator is given by

$$\begin{aligned}&\mathrm{var}\Big\{\hat{U}_{_N}(\theta_{_N}) \mid \mathcal{F}_{_N}\Big\}\\ =\ & N^{-2}\{n(n-1)\}^{-1}\Big\{\sum_{i\in\mathcal{S}} z_i^{-2}g_i(\theta_{_N})g_i(\theta_{_N})^{\mathrm{T}} - n\hat{R}(\theta_{_N})\hat{R}(\theta_{_N})^{\mathrm{T}}\Big\}\,,\end{aligned}$$

where $\hat{R}(\theta_{_N}) = n^{-1}\sum_{i\in\mathcal{S}} z_i^{-1}g_i(\theta_{_N}) = \sum_{i\in\mathcal{S}}\pi_i^{-1}g_i(\theta_{_N}) = O_p(Nn^{-1/2})$. It follows that the last term involving $\hat{R}(\theta_{_N})$ can be dropped and $\mathrm{var}\Big\{\hat{U}_{_N}(\theta_{_N}) \mid \mathcal{F}_{_N}\Big\}$ is asymptotically equivalent to $n_{_B}^{-1}W = N^{-2}\sum_{i=1}^N \pi_i^{-1}[g_i(\theta_{_N})][g_i(\theta_{_N})]^{\mathrm{T}}$. In other words, we can replace Ω by $n_{_B}^{-1}W$ in V_1, which reduces to V_3. The result is also valid for single-stage PPS sampling without replacement with negligible sampling fractions. □

Proof of Theorem 2. It follows from the proof of Theorem 1 that

$$\begin{aligned}&2n_{_B}l(\hat{\theta}_{_{SEL}},\hat{\lambda}_{_{SEL}}) = 2n_{_B}L(\tilde{\theta},\tilde{\lambda}) + o_p(1)\\ &= 2n_{_B}\{[\Gamma(\theta_{_N})(\tilde{\theta}-\theta_{_N})]^{\mathrm{T}}\tilde{\lambda} + \hat{U}_{_N}(\theta_{_N})^{\mathrm{T}}\tilde{\lambda} - \frac{1}{2}\tilde{\lambda}^{\mathrm{T}}W\tilde{\lambda}\} + o_p(1)\,.\end{aligned}$$

By the first-order condition $\Gamma(\theta_{_N})(\tilde{\theta}-\theta_{_N}) + \hat{U}_{_N}(\theta_{_N}) - W\tilde{\lambda} = 0$ from (S3.2), we have

$$2n_{_B}l(\hat{\theta}_{_{SEL}},\hat{\lambda}_{_{SEL}}) = n_{_B}\tilde{\lambda}^{\mathrm{T}}W\tilde{\lambda} + o_p(1)\,.$$

Recall that $P = W^{-1} - W^{-1}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1}$ is defined in the proof of Theorem 1 and $\tilde{\lambda} = P\hat{U}_{_N}(\theta_{_N})$. It follows from $PWP = P$ that

$$2n_{_B}l(\hat{\theta}_{_{SEL}},\hat{\lambda}_{_{SEL}}) = n_{_B}\hat{U}_{_N}(\theta_{_N})^{\mathrm{T}}P\hat{U}_{_N}(\theta_{_N}) + o_p(1)\,.$$

We also have $2n_{_B}l(\theta_{_N},\lambda)=n_{_B}\hat{U}_{_N}(\theta_{_N})^{\rm T}W^{-1}\hat{U}_{_N}(\theta_{_N})+o_p(1)$. It follows that

$$\begin{aligned} T_n(\theta_{_N}) &= 2\{l_n(\theta_{_N},\lambda)-l_n(\hat{\theta}_{_{SEL}},\hat{\lambda}_{_{SEL}})\} \\ &= n_{_B}^{1/2}\hat{U}_{_N}(\theta_{_N})^{\rm T}W^{-1}\Gamma\Sigma\Gamma^{\rm T}W^{-1}n_{_B}^{1/2}\hat{U}_{_N}(\theta_{_N})+o_p(1). \end{aligned}$$

The final result follows from the asymptotic normality of $n_{_B}^{1/2}\hat{U}_{_N}(\theta_{_N})$ with mean 0 and variance-covariance matrix $n_{_B}\Omega$. □

Proof of Theorem 3. The restricted estimators $\hat{\theta}^*_{_{SEL}}$ and $\hat{\lambda}^*_{_{SEL}}$ under H_0: $\Phi(\theta_{_N})=0$ are the optimizers of $\min_{\theta\in\Theta}\sup_{\lambda\in\mathcal{R}^q}\{l_n(\theta,\lambda)+\xi^{\rm T}\Phi(\theta)\}$, where ξ is a $k\times 1$ vector of Lagrange multipliers. Recall that $l_n(\theta,\lambda)$ can be approximated by the quadratic form

$$L(\theta,\lambda)=[\Gamma(\theta_{_N})(\theta-\theta_{_N})]^{\rm T}\lambda+\hat{U}_{_N}(\theta_{_N})^{\rm T}\lambda-0.5\lambda^{\rm T}W\lambda\,.$$

Let $\tilde{\theta}^*$ and $\tilde{\lambda}^*$ be the optimizers of $\min_{\theta\in\Theta}\sup_{\lambda\in\mathcal{R}^q}\{L(\theta,\lambda)+\xi^{\rm T}\Phi(\theta)\}$. Using similar arguments to the proof of Theorem 1, we can show that $\|\hat{\theta}^*_{_{SEL}}-\tilde{\theta}^*\|=o_p(n_{_B}^{-1/2})$ and $\|\hat{\lambda}^*_{_{SEL}}-\tilde{\lambda}^*\|=o_p(n_{_B}^{-1/2})$. Hence we only need to focus on $\tilde{\theta}^*$ and $\tilde{\lambda}^*$. The first-order conditions for an interior global maximum are given by

$$\begin{aligned} \Gamma^{\rm T}(\theta_{_N})\tilde{\lambda}^*+\Psi(\tilde{\theta}^*)^{\rm T}\tilde{\xi}^* &= 0\,, \\ \Gamma(\theta_{_N})(\tilde{\theta}^*-\theta_{_N})+\hat{U}_{_N}(\theta_{_N})-W\tilde{\lambda}^* &= 0\,, \\ \Phi(\tilde{\theta}^*) &= 0\,, \end{aligned}$$

where $\Psi(\theta)=\partial\Phi(\theta)/\partial\theta$. An identical argument to that in Step 1 of the proof of Theorem 1 shows that $\|\tilde{\theta}^*-\theta_{_N}\|=O_p(n_{_B}^{-1/2})$. We have $\Psi(\tilde{\theta}^*)=$

$\Psi(\theta_{\scriptscriptstyle N})+O_p(n_{\scriptscriptstyle B}^{-1/2})$ and $\Phi(\tilde\theta^*)=\Psi(\theta_{\scriptscriptstyle N})(\tilde\theta^*-\theta_{\scriptscriptstyle N})+o_p(n_{\scriptscriptstyle B}^{-1/2})$. It follows from $\tilde\lambda^*=O_p(n_{\scriptscriptstyle B}^{-1/2})$ that we also have $\tilde\xi^*=O_p(n_{\scriptscriptstyle B}^{-1/2})$. The first-order conditions become

$$\begin{aligned}
\Gamma(\theta_{\scriptscriptstyle N})^{\mathrm T}\tilde\lambda^*+\Psi(\theta_{\scriptscriptstyle N})^{\mathrm T}\tilde\xi^* &= o_p(n_{\scriptscriptstyle B}^{-1/2})\,,\\
\Gamma(\theta_{\scriptscriptstyle N})(\tilde\theta^*-\theta_{\scriptscriptstyle N})+\hat U_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})-W\tilde\lambda^* &= 0\,,\\
\Psi(\theta_{\scriptscriptstyle N})(\tilde\theta^*-\theta_{\scriptscriptstyle N}) &= o_p(n_{\scriptscriptstyle B}^{-1/2})\,,
\end{aligned}$$

which can be rewritten in the following matrix form

$$\begin{pmatrix} -W & \Gamma & 0\\ \Gamma^{\mathrm T} & 0 & \Psi^{\mathrm T}\\ 0 & \Psi & 0\end{pmatrix}\begin{pmatrix}\tilde\lambda^*\\ \tilde\theta^*-\theta_{\scriptscriptstyle N}\\ \tilde\xi^*\end{pmatrix}=\begin{pmatrix}-\hat U_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N})\\ 0\\ 0\end{pmatrix}+o_p(n_{\scriptscriptstyle B}^{-1/2})\,,$$

where $\Psi=\Psi(\theta_{\scriptscriptstyle N})$ and $\Gamma=\Gamma(\theta_{\scriptscriptstyle N})$. Let

$$M=\begin{pmatrix} -W & \Gamma & 0\\ \Gamma^{\mathrm T} & 0 & \Psi^{\mathrm T}\\ 0 & \Psi & 0\end{pmatrix}=:\begin{pmatrix} M_{11} & M_{12}\\ M_{21} & M_{22}\end{pmatrix}\,,$$

with $M_{11}=-W$, $M_{12}=(\Gamma,0)$, $M_{21}=M_{12}^{\mathrm T}$ and

$$M_{22}=\begin{pmatrix} 0 & \Psi^{\mathrm T}\\ \Psi & 0\end{pmatrix}\,.$$

Using results from block matrix inversions, we have

$$M^{-1}=\begin{pmatrix} M_{11}^{-1} & 0\\ 0 & 0\end{pmatrix}+\begin{pmatrix}-M_{11}^{-1}M_{12}\\ I\end{pmatrix}D^{-1}(-M_{21}M_{11}^{-1}\quad I)\,,$$

where

$$D = M_{22} - M_{21}M_{11}^{-1}M_{12} = \begin{pmatrix} \Sigma^{-1} & \Psi^{\mathrm{T}} \\ \Psi & 0 \end{pmatrix}.$$

Applying the result from block matrix inversions to D, we have

$$D^{-1} = \begin{pmatrix} \Sigma - \Sigma\Psi^{\mathrm{T}}(\Psi\Sigma\Psi^{\mathrm{T}})^{-1}\Psi\Sigma & -\Sigma\Psi^{\mathrm{T}}(\Psi\Sigma\Psi^{\mathrm{T}})^{-1} \\ -(\Psi\Sigma\Psi^{\mathrm{T}})^{-1}\Psi\Sigma & (\Psi\Sigma\Psi^{\mathrm{T}})^{-1} \end{pmatrix}.$$

It leads to

$$\begin{pmatrix} \tilde{\theta}^{*} - \theta_{N} \\ \tilde{\xi}^{*} \end{pmatrix} = D^{-1}M_{21}M_{11}^{-1}\hat{U}_{N}(\theta_{N}) + o_{p}(n_{B}^{-1/2}),$$

and $\tilde{\lambda}^{*} = -[M_{11}^{-1} + M_{11}^{-1}M_{12}D^{-1}M_{21}M_{11}^{-1}]\hat{U}_{N}(\theta_{N}) + o_{p}(n_{B}^{-1/2})$. Additionally, we have that

$$\tilde{\theta}^{*} - \theta_{N} = -P_{1}^{*}\Gamma^{\mathrm{T}}W^{-1}\hat{U}_{N}(\theta_{N}) + o_{p}(n_{B}^{-1/2}),$$

$$\tilde{\xi}^{*} = (\Psi\Sigma\Psi^{\mathrm{T}})^{-1}\Psi\Sigma\Gamma^{\mathrm{T}}W^{-1}\hat{U}_{N}(\theta_{N}) + o_{p}(n_{B}^{-1/2}),$$

$$\tilde{\lambda}^{*} = P_{2}^{*}\hat{U}_{N}(\theta_{N}) + o_{p}(n_{B}^{-1/2}),$$

where $P_{1}^{*} = \Sigma - \Sigma\Psi^{\mathrm{T}}(\Psi\Sigma\Psi^{\mathrm{T}})^{-1}\Psi\Sigma$ and $P_{2}^{*} = W^{-1} - W^{-1}\Gamma P_{1}^{*}\Gamma^{\mathrm{T}}W^{-1}$. It follows that the restricted estimator $\tilde{\theta}^{*}$ is asymptotically normal with mean θ_{N} and variance-covariance matrix $V^{*} = P_{1}^{*}\Gamma^{\mathrm{T}}W^{-1}\Omega W^{-1}\Gamma P_{1}^{*}$, which is the result for Part (i) of the theorem.

We are now ready to derive the asymptotic distribution of the sample empirical log-likelihood ratio statistic $T_{n}(\theta_{N} \mid H_{0}) = -2\{l_{n}(\hat{\theta}_{SEL}, \hat{\lambda}_{SEL}) -$

$l_n(\hat\theta^*_{SEL}, \hat\lambda^*_{SEL})\}$. Using the same arguments to the proof of Theorem 2, we can show that

$$2n_B l(\hat\theta^*_{SEL}, \hat\lambda^*_{SEL}) = n_B \tilde\lambda^{*\mathrm{T}} W \tilde\lambda^* + o_p(1)\,.$$

This, together with $P_2^* W P_2^* = P_2^*$ and $\tilde\lambda^* = P_2^* \hat U_N(\theta_N) + o_p(n_B^{-1/2})$, leads to

$$2n_B l(\hat\theta^*_{SEL}, \hat\lambda^*_{SEL}) = n_B \hat U_N(\theta_N)^{\mathrm{T}} P_2^* \hat U_N(\theta_N) + o_p(1)\,.$$

In the proof of Theorem 2, we have already obtained

$$2n_B l(\hat\theta_{SEL}, \hat\lambda_{SEL}) = n_B \hat U_N(\theta_N)^{\mathrm{T}} P \hat U_N(\theta_N) + o_p(1)\,,$$

where $P = W^{-1} - W^{-1}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1}$. Therefore,

$$T_n(\theta_N \mid H_0) = n_B^{1/2} \hat U_N(\theta_N)^{\mathrm{T}} P_3^* n_B^{1/2} \hat U_N(\theta_N) + o_p(1)\,,$$

where $P_3^* = P_2^* - P = W^{-1}\Gamma(\Sigma - P_1^*)\Gamma^{\mathrm{T}}W^{-1}$. The sample empirical likelihood ratio statistic $T_n(\theta_N \mid H_0)$ converges in distribution to $Q^{\mathrm{T}}\Delta^* Q$, where $Q \sim N(0, I_r)$ and $\Delta^* = n_B \Omega^{1/2} W^{-1}\Gamma(\Sigma - P_1^*)\Gamma^{\mathrm{T}} W^{-1}\Omega^{1/2}$. □

Proof of Corollary 2. The simplification of the result follows from using $n_B\Omega = W$, which leads to

$$n_B \Omega^{1/2} W^{-1}\Gamma\Sigma\Gamma^{\mathrm{T}} W^{-1}\Omega^{1/2} = W^{-1/2}\Gamma\Sigma\Gamma^{\mathrm{T}} W^{-1/2}\,.$$

Let $\Delta = W^{-1/2}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1/2}$. It is clear that Δ is symmetric and idempotent,

$$\mathrm{trace}(\Delta) = \mathrm{trace}\{\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1}\} = \mathrm{trace}\{\Sigma\Gamma^{\mathrm{T}}W^{-1}\Gamma\} = p\,.$$

Hence, the empirical log-likelihood ratio statistic $T_n(\theta_{\scriptscriptstyle N})$ is asymptotically distributed as $\chi^2(p)$. □

Proof of Corollary 3. We have $n_{\scriptscriptstyle B}\Omega = W$ under the given sampling designs, which leads to

$$V^* = V_3 - V_3\Psi^{\mathrm{T}}(\Psi V_3\Psi^{\mathrm{T}})^{-1}\Psi V_3\,,$$

where $V_3 = (n_{\scriptscriptstyle B}\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1}$. Furthermore, we have

$$n_{\scriptscriptstyle B}\Omega^{1/2}W^{-1}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1}\Omega^{1/2} = W^{-1/2}\Gamma\Sigma\Gamma^{\mathrm{T}}W^{-1/2} =: D_1^*\,,$$
$$n_{\scriptscriptstyle B}\Omega^{1/2}W^{-1}\Gamma P_1^*\Gamma^{\mathrm{T}}W^{-1}\Omega^{1/2} = W^{-1/2}\Gamma P_1^*\Gamma^{\mathrm{T}}W^{-1/2} =: D_2^*\,,$$

and the two matrices D_1^* and D_2^* satisfy

$$\begin{aligned}\mathrm{trace}(D_1^*) &= \mathrm{trace}\{\Sigma\Gamma^{\mathrm{T}}W^{-1}\Gamma\} = \mathrm{trace}\{I_p\} = p\,,\\ \mathrm{trace}(D_2^*) &= \mathrm{trace}\{P_1^*\Gamma^{\mathrm{T}}W^{-1}\Gamma\}\\ &= \mathrm{trace}\{I_p\} - \mathrm{trace}\{\Sigma\Psi^{\mathrm{T}}(\Psi\Sigma\Psi^{\mathrm{T}})^{-1}\Psi\} = p - k\,.\end{aligned}$$

It follows that the test statistic $T_n(\theta_{\scriptscriptstyle N} \mid H_0)$ converges in distribution to a χ^2 random variable with $p-(p-k) = k$ degrees of freedom as $N \to \infty$. □

Proof of Theorem 4. We consider the following penalized sample empirical likelihood function

$$l_{\tau_n}(\theta,\lambda) = n_{\scriptscriptstyle B}^{-1}\sum_{i\in\mathcal{S}}\log\{1+\lambda^{\mathrm{T}}\pi_i^{-1}f_{\scriptscriptstyle N}g_i(\theta)\} + \sum_{j=1}^{p}p_{\tau_n}(|\theta_j|)\,.$$

It follows from the proof of Theorem 1 that $l_{\tau_n}(\theta,\lambda)$ can be approximated by

$$L_{\tau_n}(\theta,\lambda) = [\Gamma(\theta_{_N})(\theta-\theta_{_N})]^{\mathrm{T}}\lambda + \hat{U}_{N}(\theta_{_N})^{\mathrm{T}}\lambda - \frac{1}{2}\lambda^{\mathrm{T}}W\lambda + \sum_{j=1}^{p} p_{\tau_n}(|\theta_j|)\,.$$

Using the same notation $\Gamma = \Gamma(\theta_{_N})$, we obtain that

$$\partial L_{\tau_n}(\theta,\lambda)/\partial\theta_j = \{\Gamma^{\mathrm{T}}\lambda\}_j + p'_{\tau_n}(|\theta_j|)\mathrm{sign}(\theta_j)$$

for $\theta \in \left\{\theta : \|\theta-\theta_{_N}\| \le c n_{_B}^{-1/2}\right\}$ with the jth component θ_j, where $\{\Gamma^{\mathrm{T}}\lambda\}_j$ denotes the jth component of vector $\Gamma^{\mathrm{T}}\lambda$. It can be shown that $\{\Gamma^{\mathrm{T}}\lambda\}_j = O_p(n_{_B}^{-1/2})$. Thus, for every $j \in \mathcal{A}^C$ (the set of zero coefficients), we have

$$\partial L_{\tau_n}(\theta,\lambda)/\partial\theta_j = \tau_n\left\{\tau_n^{-1}p'_{\tau_n}(|\theta_j|)\mathrm{sign}(\theta_j) + O_p(n_{_B}^{-1/2}/\tau_n)\right\}.$$

With Condition C7, we can show that $p'_{\tau_n}(|\theta_j|)\mathrm{sign}(\theta_j)$ dominates the sign of $\partial L_{\tau_n}(\theta,\lambda)/\partial\theta_j$ for all $j \in \mathcal{A}^C$. Therefore, we can show that for any $j \in \mathcal{A}^C$ and with probability tending to one,

$$\frac{\partial L_{\tau_n}(\theta,\lambda)}{\partial\theta_j} < 0 \text{ for } \theta_j \in (0,\varepsilon_n)\,, \quad \frac{\partial L_{\tau_n}(\theta,\lambda)}{\partial\theta_j} > 0 \text{ for } \theta_j \in (-\varepsilon_n, 0)$$

for any $\varepsilon_n = c n_{_B}^{-1/2}$ and a given $c > 0$. It follows from the arguments of Fan and Li (2001) that, with probability approaching to 1, $\hat{\theta}_j = 0$ for all $j \in \mathcal{A}^C$. Recall that $\theta_{_N} = (\theta_{_{N1}}^{\mathrm{T}}, \theta_{_{N2}}^{\mathrm{T}})^{\mathrm{T}}$, where $\theta_{_{N2}} = 0$, and $\hat{\theta}_{_{PSEL}} = (\hat{\theta}_{_{P1}}^{\mathrm{T}}, \hat{\theta}_{_{P2}}^{\mathrm{T}})^{\mathrm{T}}$. We have proved Part (i) of the theorem that $\mathrm{P}(\hat{\theta}_{_{P2}} = 0 \mid \mathcal{F}_{_N}) \to 1$ as $N \to \infty$.

Let H_1 and H_2 be the two matrices such that $H_1\theta_{_N} = \theta_{_{N1}}$ and $H_2\theta_{_N} = \theta_{_{N2}}$. Estimation of $\theta_{_{N1}}$ is equivalent to estimation of $\theta_{_N}$ under the constraints

$H_2\theta_{_N} = 0$. Following the same techniques used in the proof of Theorem 3, which were previously used by Qin and Lawless (1995) and Tang and Leng (2010), we see that finding the minimizer of $L_{\tau_n}(\theta, \lambda)$ is asymptotically equivalent to solving the minimization of the objective function

$$[\Gamma(\theta_{_N})(\theta - \theta_{_N})]^{\mathrm{T}}\lambda + \hat{U}_{_N}(\theta_{_N})^{\mathrm{T}}\lambda - \frac{1}{2}\lambda^{\mathrm{T}}W\lambda + \sum_{j=1}^{p} p_{\tau_n}(|\theta_j|) + \nu^{\mathrm{T}}H_2\theta\,,$$

where ν is a $(p-d)\times 1$ vector of Lagrange multipliers. Without loss of generality, we denote the minimizer of above objective function by $(\hat{\theta}_{_{PSEL}}, \hat{\lambda}_{_{PSEL}}, \hat{\nu}_{_{PSEL}})$. The first-order conditions for the global minimizer are given by

$$\begin{pmatrix} \hat{\lambda}_{_{PSEL}} \\ \hat{\theta}_{_{PSEL}} - \theta_{_N} \\ \hat{\nu}_{_{PSEL}} \end{pmatrix} = K^{-1} \begin{pmatrix} -\hat{U}_{_N}(\theta_{_N}) \\ 0 \\ 0 \end{pmatrix} + o_p(n_{_B}^{-1/2})\,,$$

where

$$K = \begin{pmatrix} -W & \Gamma & 0 \\ \Gamma^{\mathrm{T}} & 0 & H_2^{\mathrm{T}} \\ 0 & H_2 & 0 \end{pmatrix}.$$

Using similar arguments to those given in the proof of Theorem 3, we can show that

$$\hat{\theta}_{_{PSEL}} - \theta_{_N} = -\{I - \Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\}\Sigma\Gamma^{\mathrm{T}}W^{-1}\hat{U}_{_N}(\theta_{_N}) + o_p(n_{_B}^{-1/2})$$

and

$$\hat{\theta}_{_{P1}} - \theta_{_{N1}} = -\{H_1 - H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\}\Sigma\Gamma^{\mathrm{T}}W^{-1}\hat{U}_{_N}(\theta_{_N}) + o_p(n_{_B}^{-1/2})\,.$$

Asymptotic normality of $\hat{\theta}_{P1}$ follows immediately with mean θ_{N1} and variance-covariance matrix given by

$$V_{1\mathcal{A}} = \{H_1 - H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\}V_1\{H_1^{\mathrm{T}} - H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma H_1^{\mathrm{T}}\}\,,$$

where $V_1 = \Sigma\Gamma^{\mathrm{T}}W^{-1}\Omega W^{-1}\Gamma\Sigma$ is given in Theorem 1. Simple algebraic manipulations show that

$$\begin{aligned} V_{1\mathcal{A}} &= H_1V_1H_1^{\mathrm{T}} - H_1V_1H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma H_1^{\mathrm{T}} \\ &\quad -H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2V_1H_1^{\mathrm{T}} \\ &\quad +H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2V_1H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma H_1^{\mathrm{T}} \\ &= V_{111} - V_{112}\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{-1}V_{121} + \Sigma_{12}\Sigma_{22}^{-1}V_{122}\Sigma_{22}^{-1}\Sigma_{21}\,. \end{aligned}$$

This completes the proof of Theorem 4. □

Proof of Theorem 5. Recall that the penalized sample empirical likelihood is defined as

$$l_{\tau_n}(\theta,\lambda) = n_B^{-1}\sum_{i\in\mathcal{S}}\log\{1+\lambda^{\mathrm{T}}\pi_i^{-1}f_N g_i(\theta)\} + \sum_{j=1}^{p} p_{\tau_n}(|\theta_j|)\,.$$

It follows from the proof of Theorem 1 that

$$\begin{aligned}
2n_{B}l_{\tau_n}(\hat{\theta}_{PSEL},\hat{\lambda}_{PSEL}) &= 2n_{B}L(\hat{\theta}_{PSEL},\hat{\lambda}_{PSEL})+2n_{B}\sum_{j=1}^{p}p_{\tau_n}(|\hat{\theta}_j|)+o_p(1)\\
&= 2n_{B}\Big\{[\Gamma(\theta_{N})(\hat{\theta}_{PSEL}-\theta_{N})]^{\mathrm{T}}\hat{\lambda}_{PSEL}+\hat{U}_{N}(\theta_{N})^{\mathrm{T}}\hat{\lambda}_{PSEL}\\
&\quad -\frac{1}{2}\hat{\lambda}_{PSEL}^{\mathrm{T}}W\hat{\lambda}_{PSEL}\Big\}+2n_{B}\sum_{j=1}^{p}p_{\tau_n}(|\hat{\theta}_j|)+o_p(1)\\
&= n_{B}\hat{\lambda}_{PSEL}^{\mathrm{T}}W\hat{\lambda}_{PSEL}+2n_{B}\sum_{j=1}^{p}p_{\tau_n}(|\hat{\theta}_j|)+o_p(1)\,,
\end{aligned}$$

where $\hat{\theta}_j$ is the j-th component of $\hat{\theta}_{PSEL}$. From the proof of Theorem 2 and for the penalized sample empirical likelihood method, we have

$$\hat{\lambda}_{PSEL}=\{W^{-1}-W^{-1}\Gamma\dot{Z}_{11}\Gamma^{\mathrm{T}}W^{-1}\}\hat{U}_{N}(\theta_{N})+o_p(n_{B}^{-1/2}),$$

where $\dot{Z}_{11}=\Sigma-\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma$. Note that there exists $\tilde{H}_2$ such that $\tilde{H}_2\theta_{N}=\theta_{N2}$ and $\tilde{H}_2\tilde{H}_2^{\mathrm{T}}=I_{p-d+q}$. Denote $\check{\theta}_{PSEL}=\min_{B\theta_1=0}L_p(\theta,\lambda)$. Similar to the proof of Theorem 3, we obtain that under $H_0:B\theta_{N1}=0$,

$$2n_{B}l_p(\check{\theta}_{PSEL},\check{\lambda}_{PSEL})=n_{B}\check{\lambda}_{PSEL}^{\mathrm{T}}W\check{\lambda}_{PSEL}+2n_{B}\sum_{j=1}^{p}p_{\tau_n}(|\check{\theta}_j|)+o_p(1),$$

where

$$\begin{aligned}
\check{\lambda}_{PSEL} &= \{W^{-1}-W^{-1}\Gamma\ddot{Z}_{11}\Gamma^{\mathrm{T}}W^{-1}\}\hat{U}_{N}(\theta_{N})+o_p(n_{B}^{-1/2})\,,\\
\ddot{Z}_{11} &= \Sigma-\Sigma\tilde{H}_2^{\mathrm{T}}(\tilde{H}_2\Sigma\tilde{H}_2^{\mathrm{T}})^{-1}\tilde{H}_2\Sigma\,,
\end{aligned}$$

and $\check{\theta}_j$ is the j-th component of $\check{\theta}_{PSEL}$. On the other hand, it follows from

Condition C7 and the oracle property of $\hat\theta_{PSEL}$ that

$$n_B \sum_{j=1}^{p} \{ p_{\tau_n}(|\hat\theta_j|) - p_{\tau_n}(|\check\theta_j|) \} = o_p(1)$$

as $n_B \to \infty$. Define $\dot P = \{W^{-1} - W^{-1}\Gamma \dot Z_{11} \Gamma^{\mathrm{T}} W^{-1}\}$ and $\ddot P = \{W^{-1} - W^{-1}\Gamma \ddot Z_{11} \Gamma^{\mathrm{T}} W^{-1}\}$. It can be shown that (i) $\dot Z_{11}$, $\ddot Z_{11}$, $\dot P$ and $\ddot P$ are symmetric; and (ii) $\dot Z_{11}\Sigma^{-1}\dot Z_{11} = \dot Z_{11}$, $\ddot Z_{11}\Sigma^{-1}\ddot Z_{11} = \ddot Z_{11}$, $\dot P W \dot P = \dot P$ and $\ddot P W \ddot P = \ddot P$. Combining above arguments, we have

$$T_{\tau_n}(\theta_{N1} \mid H_0) = \{n_B^{1/2} W^{-1/2} \hat U_N(\theta_N)\}^{\mathrm{T}} \{\tilde{\mathcal{P}} - \mathcal{P}\} \{n_B^{1/2} W^{-1/2} \hat U_N(\theta_N)\} + o_p(1),$$

where

$$\begin{aligned} \tilde{\mathcal{P}} &= W^{1/2}\Sigma \tilde H_2^{\mathrm{T}} (\tilde H_2 \Sigma \tilde H_2^{\mathrm{T}})^{-1} \tilde H_2 \Sigma W^{1/2}, \\ \mathcal{P} &= W^{1/2}\Sigma H_2^{\mathrm{T}} (H_2 \Sigma H_2^{\mathrm{T}})^{-1} H_2 \Sigma W^{1/2}. \end{aligned}$$

Both matrices $\tilde{\mathcal{P}}$ and P are idempotent. Since the matrix $\tilde{\mathcal{P}} - \mathcal{P}$ is also idempotent with rank$(\tilde{\mathcal{P}} - \mathcal{P}) = q$, the result of Theorem 5 follows immediately. □

Proof of Corollary 4. If $r = p$, then Γ and W are both invertible $p \times p$

matrices and V_1 reduces to $V_2 = (\Gamma^{\mathrm{T}}\Omega^{-1}\Gamma)^{-1} = \Gamma^{-1}\Omega(\Gamma^{\mathrm{T}})^{-1}$. We have

$$\begin{aligned}
V_{1\mathcal{A}} &= H_1V_2H_1^{\mathrm{T}} - H_1V_2H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma H_1^{\mathrm{T}} \\
&\quad -H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2V_2H_1^{\mathrm{T}} \\
&\quad +H_1\Sigma H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2V_2H_2^{\mathrm{T}}(H_2\Sigma H_2^{\mathrm{T}})^{-1}H_2\Sigma H_1^{\mathrm{T}} \\
&= V_{211} - V_{212}\Sigma_{22}^{-1}\Sigma_{21} - \Sigma_{12}\Sigma_{22}^{-1}V_{221} + \Sigma_{12}\Sigma_{22}^{-1}V_{222}\Sigma_{22}^{-1}\Sigma_{21} \\
&=: V_{2\mathcal{A}}\,.
\end{aligned}$$

Part (i) of Corollary 4 then follows.

Under single-stage PPS sampling with replacement, we have $\Omega = n_{B}^{-1}W$, and V_1 reduces to $V_3 = (n_B\Gamma^{\mathrm{T}}W^{-1}\Gamma)^{-1} = n_{B}^{-1}\Sigma$. Then

$$\begin{aligned}
V_{1\mathcal{A}} &= H_1V_3H_1^{\mathrm{T}} - H_1V_3H_2^{\mathrm{T}}(H_2V_3H_2^{\mathrm{T}})^{-1}H_2V_3H_1^{\mathrm{T}} \\
&\quad -H_1V_3H_2^{\mathrm{T}}(H_2V_3H_2^{\mathrm{T}})^{-1}H_2V_3H_1^{\mathrm{T}} \\
&\quad +H_1V_3H_2^{\mathrm{T}}(H_2V_3H_2^{\mathrm{T}})^{-1}H_2V_3H_2^{\mathrm{T}}(H_2V_3H_2^{\mathrm{T}})^{-1}H_2V_3H_1^{\mathrm{T}} \\
&= H_1V_3H_1^{\mathrm{T}} - H_1V_3H_2^{\mathrm{T}}(H_2V_3H_2^{\mathrm{T}})^{-1}H_2V_3H_1^{\mathrm{T}} \\
&= V_{311} - V_{312}V_{322}^{-1}V_{321} \\
&=: V_{3\mathcal{A}}\,.
\end{aligned}$$

This completes the proof of Corollary 4. □

S4 Practical Implementation on Variance Estimation

One of the most crucial practical implementations for the proposed methods is the estimation of the third quantity $\Omega = \mathrm{Var}[\hat{U}_{\scriptscriptstyle N}(\theta_{\scriptscriptstyle N}) \mid \mathcal{F}_{\scriptscriptstyle N}]$, which amounts to design-based variance estimation for the Horvitz-Thompson estimator. This is one of the major topics in survey sampling and is not unique to the sample empirical likelihood methods developed in this paper. For single-stage PPS sampling without replacement with small sampling fractions, the results presented in Sections 2 and 3 do not require the estimation of Ω. We provide details for three other commonly encountered sampling designs in survey practice. Each of these designs was examined in the simulation studies.

S4.1 Single-stage PPS sampling with non-negligible sampling fractions

The challenge for design-based variance estimation is the requirement on second order inclusion probabilities $\pi_{ij} = \mathrm{P}(i, j \in \mathcal{S})$, which are typically unavailable to survey data users. For single-stage PPS sampling with non-negligible sampling fractions, there exist approximate variance formulas which do not involve the π_{ij}; see Haziza et al. (2008) for a review of the topic with relevant references. For high entropy sampling designs such as the well-known Rao-Sampford method, we can estimate Ω by

$$\hat{\Omega} = \frac{1}{N^2} \sum_{i \in \mathcal{S}} c_i \big[\pi_i^{-1} g_i(\hat{\theta}_{SEL}) - \hat{B}\big] \big[\pi_i^{-1} g_i(\hat{\theta}_{SEL}) - \hat{B}\big]^{\mathrm{T}} ,$$

where

$$\hat{B} = \Big\{ \sum_{i \in \mathcal{S}} c_i \pi_i^{-1} g_i(\hat{\theta}_{SEL}) \Big\} / \Big\{ \sum_{i \in \mathcal{S}} c_i \Big\} \quad \text{and} \quad c_i = \{n(1-\pi_i)\}/\{n-1\} .$$

Haziza et al. (2008) showed through simulation studies that the approximate variance estimator has excellent performance even for small sample sizes.

S4.2 Stratified sampling

Suppose that the finite population $\mathcal{U}$ is divided into H strata with stratum population sizes $N_1, N_2, \cdots, N_H$. Let $\mathcal{S}_h$ be the sample of size n_h selected from the hth stratum with inclusion probabilities $\{\pi_{hi}, i \in \mathcal{S}_h\}$ for $h = 1, \cdots, H$, independent among different strata. Let $n = \sum_{h=1}^{H} n_h$ be the size of the overall stratified sample.

Empirical likelihood methods for stratified survey samples can take two different forms. The first treats the stratified survey sample as a multiple sample problem, with the nonparametric probability measure $(p_{h1}, \cdots, p_{hn_h})$ normalized for each stratum sample, i.e., $\sum_{i \in \mathcal{S}_h} p_{hi} = 1$, $h = 1, \cdots, H$ as in Berger and Torres (2016). We propose to use the second form, the pooled sample approach where the sample empirical log-likelihood is computed and the constraints are formulated by treating the stratified sample as a single pooled sample, with the single normalization constraint $\sum_{h=1}^{H}\sum_{i\in\mathcal{S}_h} p_{hi} = 1$. The pooled sample approach provides a unified computational framework for both stratified and non-stratified sampling, since all procedures for non-stratified sampling can be used directly for stratified sampling. The stratified sampling design feature is taken into account through the estimation of Ω, which takes the form

$$\Omega = \frac{1}{N^2}\sum_{h=1}^{H}\mathrm{Var}\Big\{\sum_{i\in\mathcal{S}_h}\pi_{hi}^{-1}g_{hi}(\theta_{\scriptscriptstyle N}) \mid \mathcal{F}_{\scriptscriptstyle N}\Big\}\,.$$

If the stratum samples $\mathcal{S}_h$ are selected by a PPS sampling design with small sampling fractions, we can estimate Ω by

$$\hat{\Omega} = \frac{1}{N^2}\sum_{h=1}^{H}\sum_{i\in\mathcal{S}_h}\Big[\pi_{hi}^{-1}g_{hi}(\hat{\theta}_{SEL}) - \hat{U}_h(\hat{\theta}_{SEL})\Big]\Big[\pi_{hi}^{-1}g_{hi}(\hat{\theta}_{SEL}) - \hat{U}_h(\hat{\theta}_{SEL})\Big]^{\mathrm{T}},$$

where $\hat{U}_h(\theta) = (1/n_h)\sum_{i\in\mathcal{S}_h}\pi_{hi}^{-1}g_{hi}(X_{hi},Y_{hi},\theta)$. If the sampling fraction is not small, approximate variance formulas can be used for each stratum sample to get an estimate for the overall variance Ω.

S4.3 Cluster sampling

We consider two-stage cluster sampling designs where the first stage clusters are selected by a PPS sampling method, with inclusion probabilities proportional to cluster sizes, and the second stage units are selected by simple random sampling without replacement. Let K be the total number of clusters in the population and $\{M_1, M_2, \cdots, M_K\}$ be the cluster sizes. The population size is given by $N = \sum_{i=1}^{K} M_i$.

Let $\mathcal{S}_c$ be the set of k clusters selected from the population, with inclusion probabilities $\pi_{1i} = \mathrm{P}(i \in \mathcal{S}_c)$ proportional to M_i, i.e., $\pi_{1i} = kM_i/N$, where the subscript "$1i$" represents "first stage unit i". Let $\mathcal{S}_i$ be the set of m ($\leq M_i$) second-stage units selected from cluster i, $i \in \mathcal{S}_c$, using simple random sampling without replacement. The second-stage inclusion probabilities are given by $\pi_{j|i} = \mathrm{P}(j \in \mathcal{S}_i \mid i \in \mathcal{S}_c) = m/M_i$. The final first order inclusion probability for unit (ij) is given by $\pi_{(ij)} = \mathrm{P}(i \in \mathcal{S}_c, j \in \mathcal{S}_i) = km/N$. This is the well-known two-stage sampling design which leads to *self-weighting* for the Horvitz-Thompson estimator.

Let $g_{(ij)}(\theta)$ denote the estimating function for unit (ij). The task under the two-stage survey design is to estimate

$$\Omega = \mathrm{Var}\Big\{(km)^{-1} \sum_{i \in \mathcal{S}_c} \sum_{j \in \mathcal{S}_i} g_{(ij)}(\theta_{\scriptscriptstyle N}) \mid \mathcal{F}_{\scriptscriptstyle N}\Big\}$$

. Let $\bar{G}_i = m^{-1} \sum_{i \in \mathcal{S}_i} g_{(ij)}(\hat{\theta}_{\scriptscriptstyle SEL})$, $\bar{G} = (km)^{-1} \sum_{i \in \mathcal{S}_c} \sum_{j \in \mathcal{S}_i} g_{(ij)}(\hat{\theta}_{\scriptscriptstyle SEL})$. A design-unbiased estimator of Ω is given by

$$\hat{\Omega} = \frac{1}{k(k-1)} \sum_{i \in \mathcal{S}_c} \left[\bar{G}_i - \bar{G}\right]\left[\bar{G}_i - \bar{G}\right]^{\mathrm{T}}.$$

The results described in (1), (2) and (3) can be used to deal with stratified multi-stage sampling to obtain an estimate for Ω under such designs.

S5 Computational Issues

It is a rather challenging task to maximize the sample empirical likelihood function under a set of constraints when the estimating functions are not smooth and the gradient methods are not well defined. The development of certain derivative-free optimization algorithms becomes necessary. In this paper, we apply the idea of the Nelder-Mead (NM) simplex algorithm (Nelder and Mead (1965)) to both the un-penalized and penalized sample empirical likelihood functions.

We first use the method described in Owen (2001) to overcome the bounded support problem of the logarithm function through the following pseudo-logarithm function

$$\log_*(x) = \begin{cases} \log(x)\,, & \text{if} \quad x \geq \zeta\,, \\ \log(\zeta) - 1.5 + 2x/\zeta - x^2/(2\zeta^2)\,, & \text{if} \quad x < \zeta\,, \end{cases}$$

where ζ is usually chosen to be n^{-1}. For the maximization of the sample empirical likelihood, we use $l_n^*(\theta) = \sum_{i\in\mathcal{S}} \log_*\{1 + \lambda^{\mathrm{T}}\pi_i^{-1}g_i(\theta)\}$ as a surrogate for $l_n(\theta)$, which is computationally more stable than the original sample empirical likelihood.

For the penalized sample empirical likelihood estimator, we use the local quadratic form of Fan and Li (2001) to approximate the penalty term $p_{\tau_n}(|\theta_j|)$, which is given by

$$\tilde{p}_{\tau_n}(|\theta_j|) = p_{\tau_n}(|\theta_j^{(m)}|) + \frac{1}{2}\{p_{\tau_n}(|\theta_j^{(m)}|)/|\theta_j^{(m)}|\}\{\theta_j^2 - (\theta_j^{(m)})^2\}\,,$$

where $\theta_j^{(m)}$ is the mth step estimate of θ_j from the previous iteration.

The main NM algorithm for maximizing the sample empirical likelihood consists of an inner loop and an outer loop for the iterative procedures.

Inner Loop. The “inner loop” solves $\hat{\lambda} = \arg\min_\lambda R_*(\theta, \lambda)$ for the given θ, where $R_*(\theta, \lambda) = -\sum_{i\in\mathcal{S}} \log_*\{1 + \lambda^{\mathrm{T}}\pi_i^{-1} g_i(\theta)\}$. Since the pseudo-logarithm function is twice differentiable, such an optimization problem can be solved easily by using the modified Newton-Raphson procedure of Chen et al. (2002).

Outer Loop. Once a solution $\hat{\lambda}$ is found in the inner loop, we use the classical NM algorithm as the “outer loop” to obtain $\hat{\theta}_{SEL} = \arg\max_\theta R_*(\theta, \hat{\lambda})$ and $\hat{\theta}_{PSEL} = \arg\max_\theta \{R_*(\theta, \hat{\lambda}) - n\sum_{j=1}^p \tilde{p}_{\tau_n}(|\theta_j|)\}$.

The tuning parameter τ_n for the penalized sample empirical likelihood needs to be appropriately selected by a data-driven method. Various techniques have been proposed in the literature, including the generalized cross-validation method and the BIC method (Wang et al. (2007)). In our simulation studies, we choose the optimal value for the penalty parameter τ_n

by minimizing the following BIC-type criterion function

$$\mathrm{BIC}(\tau_n) = 2l_n(\theta, \lambda) + \log(n) df_{\tau_n}/n \,, \tag{S5.1}$$

where df_{τ_n} is the number of nonzero coefficients in the fitted model using the penalized sample empirical likelihood for the given τ_n, and $l_n(\theta)$ is the un-penalized sample empirical likelihood function.

S6 Further Details on Simulation Settings

We considered design-based inferences where the finite population was generated from a superpopulation and was fixed for repeated simulation samples. We considered four sampling designs: (I) Single-stage PPS sampling without replacement with negligible sampling fractions; (II) Single-stage PPS sampling without replacement with non-negligible sampling fractions; (III) Stratified PPS sampling; (IV) Two-stage cluster sampling with self-weighting designs. The finite population size and sample sizes were set as follows for the four sampling designs:

(I) Repeated simulation samples of size $n = 300$ were selected from the finite population of size $N = 20,000$ by the randomized systematic PPS sampling method (Hartley and Rao (1962)). This corresponds to a sampling fraction of 1.5%, which can be viewed as negligible.

(II) Repeated simulation samples of size $n = 300$ were selected from the finite population of size $N = 6,000$ by the Rao-Sampford PPS sampling method (Rao (1965); Sampford (1967)). This corresponds to a sampling fraction of 5%, which is not negligible. The variance approximation method described in the Supplementary Materials can be used for this scenario with the Rao-Sampford sampling method.

(III) The stratified finite population of size $N = 20,000$ consisted $H = 3$ strata with stratum sizes $N_1 = 4,000$, $N_2 = 6,000$ and $N_3 = 10,000$. The stratum sample sizes were set as $n_1 = 50$, $n_2 = 100$ and $n_3 = 150$. Repeated simulation samples were selected by the Rao-Sampford PPS sampling method within each stratum.

(IV) The finite population consisted of $K = 1400$ clusters, with 200 clusters having size $M_i = 30$, 400 clusters having size $M_i = 15$, and 800 clusters having size $M_i = 10$. The first stage sample $\mathcal{S}_c$ consisted $k = 60$ clusters, selected by the Rao-Sampford PPS sampling with $\pi_{1i} \propto M_i$. Within each selected cluster, a second-stage sample of size $m = 5$ is selected by simple random sampling without replacement, independent among different clusters.

S7 Additional Simulation Results

We present some additional simulation results for linear and quantile regression models. Details of the simulation settings are described in Section 6.

S7.1 Point estimation and hypothesis tests for linear regression models

The finite population follows the linear regression model

$$Y = \theta_0 + Z_1\theta_1 + Z_2\theta_2 + \sigma(Z_1, Z_2)\varepsilon\,,$$

where $(\theta_0, \theta_1, \theta_2) = (1, 1, 0.5)$, $Z_1 \sim$ Bernoulli(0.5), $Z_2 = Z_1 + Z_0$ with $Z_0 \sim N(0, 1)$, $\varepsilon \sim N(0, 1)$, and $\sigma(Z_1, Z_2)$ represents the variance structure for the error terms. We consider three variance structures for the superpopulation regression model: $\sigma_1 = \sigma(Z_1, Z_2) = 1$, $\sigma_2 = \sigma(Z_1, Z_2) = 3$ and $\sigma_3 = \sigma(Z_1, Z_2) = [Var(\eta)(1/\rho^2 - 1)]^{1/2}$ with $\eta = \theta_0 + Z_1\theta_1 + Z_2\theta_2$ and $\rho = 0.7$ which is the controlled correlation between the linear predictor η and the response variable Y.

Let $\theta = (\theta_0, \theta_1, \theta_2)^{\mathrm{T}}$, $X = (1, Z_1, Z_2)^{\mathrm{T}}$, and $X_i = (1, Z_{1i}, Z_{2i})^{\mathrm{T}}$ for $i = 1, \cdots, N$. The parameters of interest are the finite population regression coefficients $\theta_{\scriptscriptstyle N} = (\theta_{\scriptscriptstyle N0}, \theta_{\scriptscriptstyle N1}, \theta_{\scriptscriptstyle N2})^{\mathrm{T}}$ defined through the estimating functions $g(X, Y, \theta) = X(Y - X^{\mathrm{T}}\theta)$ and the census estimating equations $(1/N)\sum_{i=1}^{N} g(X_i, Y_i, \theta_{N}) = 0$. We consider finite sample performances of the maximum sample empirical likelihood estimator and the sample empirical likelihood ratio test for H_0: $\theta_{N1} = 1$ against H_1: $\theta_{N1} = b$ at the significance level 0.05.

Simulation results based on 2000 simulated samples are presented in Tables 1 and 2, where Table 1 reports the simulated relative bias (Bias) and root mean squared error (RMS) and Table 2 reports the size of power of the sample empirical likelihood ratio test for H_0: $\theta_{N1} = 1$ against H_1: $\theta_{N1} = b$. The value $b = 1$ corresponds to the size of the test and values $b \neq 1$ present the power of the test. The simulation studies are conducted for four different sampling designs I, II, III and IV as described in the main paper. It can be seen that both the point estimators and the test of our proposed sample empirical likelihood method perform well under all scenarios considered in the simulation. The power of the test is dramatically stronger when the correlations between the predictors and the response variable are higher (scenarios corresponding to σ_1 and σ_3).

S7.2 Variable selection for linear regression models

The finite population follows the linear regression model

$$Y = X^{\mathrm{T}}\theta + \sigma(X)\varepsilon,$$

where $\theta = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^{\mathrm{T}}$, the marginal distributions of $X = (X_1, \ldots, X_p)$ are standard normal with pairwise correlations $\mathrm{Corr}(X_j, X_k) = 0.5^{|j-k|}$ for $1 \leq j, k \leq p$, and $\varepsilon \sim N(0, 1)$. We consider three variance structures for the superpopulation regression model: $\sigma_1 = \sigma(X) = 1$, $\sigma_2 = \sigma(X) = 3$ and $\sigma_3 = \sigma(X) = 1 + X_2$, corresponding to three different finite populations. We consider $p = 8$ and 16 such that the number of covariates with zero coefficients is 3 and 13 respectively.

Table 3 presents the results on variable selection for linear regression models through our proposed penalized sample empirical likelihood method based on 200 simulated samples for three different variable structures under four different sampling designs. Table 3 shows that our proposed penalized sample empirical likelihood procedure for variable selection has superb performance in identifying the zero coefficients and the correct models. The results on variable selection are more accurate for linear regression models than for quantile regression models, as shown by Tables 7-9 reported in the next section.

S7.3 Variable selection for quantile regression models

We present results from the simulation study on variable selection for quantile regression models under the four sampling designs I-IV described in Section 5.1 of the main paper. The finite population was generated from the model $Y = X^{\mathrm{T}}\theta + \varepsilon(\gamma)$, where $X = (X_1, \ldots, X_p)^{\mathrm{T}}$. The marginal distributions of X are standard normal and the pairwise correlations follow $\mathrm{Corr}(X_j, X_k) = 0.5^{|j-k|}$ for $1 \le j, k \le p$. We examined five scenarios for the error terms used in the model. For the first four scenarios, we had $\varepsilon(\gamma) = \varepsilon - Q_\varepsilon(\gamma)$ with (A) $\varepsilon \sim N(0,1)$; (B) $\varepsilon \sim 0.95N(0,1) + 0.05N(0,6^2)$, a mixture of two normal distributions; (C) $\varepsilon \sim \chi^2(3)$; and (D) $\varepsilon \sim t(3)$. For the fifth scenario (E), we had $\varepsilon(\gamma) = (1 + X_1)(\varepsilon - Q_\varepsilon(\gamma))$ with $\varepsilon \sim N(0,1)$.

We considered $p = 8$ and 16, with the number of zero coefficients being 5 for $p = 8$ and 13 for $p = 16$. The true values of the coefficients were set as $\theta = (3, 1.5, 0, 0, 2, 0, \cdots, 0)^{\mathrm{T}}$. Tables 7-9 present the results on variable selection for quantile regression models with $\gamma = 0.25$, 0.50 and 0.75 using our proposed penalized sample empirical likelihood method. The results were based on $B = 200$ simulated samples. The column MSE was computed as $\mathrm{MSE}(\hat{\theta}_{PSEL}) = B^{-1}\sum_{b=1}^{B}(\hat{\theta}_{PSEL}^{(b)} - \theta_N)^{\mathrm{T}}\big(N^{-1}\sum_{i=1}^{N} X_i X_i^{\mathrm{T}}\big)(\hat{\theta}_{PSEL}^{(b)} - \theta_N)$, where $\hat{\theta}_{PSEL}^{(b)}$ was the penalized maximum sample empirical likelihood estimator $\hat{\theta}_{PSEL}$ from the bth simulated sample, $b = 1, \cdots, B$. The column labels “C” and “IC” under “No of Zeros” represent “The average number of correctly identified zeros” and “The average number of incorrectly identified zeros” for the regression coefficients, respectively. The other three columns labels “U-fit”, “C-fit” and “O-fit” under “Fitted Models” indicate the percentages of models in the simulation which were “Under-fit” (at least one non-zero coefficient was identified as zero), “Correct-fit” (all zeros and non-zeros were correctly identified), and “Over-fit” (at least one zero coefficient was selected as non-zero). Results for $\gamma = 0.50$ and 0.75 were reported in the Supplementary Materials.

The results for all three quantile regression models ($\gamma = 0.25$, 0.50 and 0.75) show that the proportions of correct-fit models are very high and the proportions of under-fit models are very close to zero. For $p = 16$, there were cases (about $10 - 15\%$) where the model was over-fit, with most of these cases having one zero coefficient selected as non-zero. Overall, the average numbers of estimated zero coefficients are close to the true value 5 for $p = 8$ and 13 for $p = 16$. The sampling designs do not seem to have significant impact on the performance of the proposed variable selection method, and the proposed method seems also to be robust towards different error distributions for the quantile regression models.

References

Berger, Y. G. and De La Riva Torres, O. (2016). Empirical Likelihood Confidence Intervals for Complex Sampling Designs. *Journal of the Royal Statistical Society: Series B*, **78**, 319–341.

Chen, J., Sitter, R. R. and Wu, C. (2002). Using Empirical Likelihood Methods to Obtain Range Restricted Weights in Regression Estimators for Surveys. *Biometrika*, **89**, 230–237.

Chen, S. and Kim, J. K. (2014). Population Empirical Likelihood for Nonparametric Inference in Survey Sampling. *Statistica Sinica*, **24**, 335–355.

Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American statistical Association*, **96**, 1348–1360.

Francisco, C. A. and Fuller, W. A. (1991). Quantile Estimation with a Complex Survey Design. *The Annals of Statistics*, **19**, 454–469.

Hartley, H. O. and Rao, J. N. K. (1962). Sampling with Unequal Probabilities and Without Replacement. *Annals of Mathematical Statistics*, **33**, 350–374.

Haziza, D., Mecatti, F. and Rao, J. N. K. (2008). Evaluation of Some Approximate Variance Estimators under the Rao-Sampford Unequal Probability Sampling Design. *Metron*, **66**, 91–108.

Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, **7**, 308–313.

Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, New York.

Oguz-Alper, M. and Berger, Y. G. (2016). Modelling Complex Survey Data with Population Level Information: An Empirical Likelihood Approach. *Biometrika*, **103**, 447–459.

Qin, J. and Lawless, J. (1995). Estimating Equations, Empirical Likelihood and Constraints on Parameters. *The Canadian Journal of Statistics*, **23**, 145–159.

Rao, J. N. K. (1965). On Two Simple Schemes of Unequal Probability Sampling Without Replacement. *Journal of the Indian Statistical Association*, **3**, 173–180.

Sampford, M. R. (1967). On Sampling Without Replacement with Unequal Probabilities of Selection. *Biometrika*, **54**, 499–513.

Tang, C. Y. and Leng, C. (2010). Penalized High Dimensional Empirical Likelihood. *Biometrika*, **97**, 905–920.

van der Vaart, A.W. and Wellner, J.A. (1996). Weak Convergence and Empirical Processes. Springer-Verlag, New York.

Wang, H., Li, R. and Tsai, C. L. (2007). Tuning Parameter Selectors for the Smoothly Clipped Absolute Deviation Method. *Biometrika*, **94**, 553–568.

Wang, J. C. and Opsomer, J. D. (2011). On Asymptotic Normality and Variance Estimation for Nondifferentiable Survey Estimators. *Biometrika*, 98, 91–106.

Table 1: Relative Bias and RMS of Point Estimators for Linear Regression

		σ_1			σ_2			σ_3		
Design		θ_{N0}	θ_{N1}	θ_{N2}	θ_{N0}	θ_{N1}	θ_{N2}	θ_{N0}	θ_{N1}	θ_{N2}
I	Bias	0.008	-0.004	-0.003	-0.005	0.015	-0.003	-0.002	0.006	-0.003
	RMS	0.104	0.136	0.065	0.285	0.423	0.197	0.093	0.128	0.057
II	Bias	0.006	-0.004	-0.002	-0.003	0.005	0.005	0.000	0.002	-0.003
	RMS	0.107	0.141	0.063	0.320	0.428	0.195	0.095	0.123	0.060
III	Bias	0.004	-0.001	-0.003	0.011	-0.005	-0.003	0.002	-0.004	0.002
	RMS	0.104	0.138	0.064	0.285	0.407	0.184	0.094	0.127	0.058
IV	Bias	-0.001	-0.001	0.000	0.006	0.003	-0.008	0.000	-0.002	-0.001
	RMS	0.082	0.129	0.058	0.243	0.380	0.174	0.074	0.118	0.052

Table 2: The Size and Power of the SEL Ratio Test for Linear Regression

Design	$\sigma(X)$	$b = 0.50$	0.75	1.00	1.25	1.50
I	σ_1	0.951	0.476	0.068	0.470	0.962
	σ_2	0.235	0.100	0.058	0.108	0.247
	σ_3	0.999	0.634	0.058	0.468	0.902
II	σ_1	0.953	0.471	0.074	0.471	0.955
	σ_2	0.258	0.121	0.071	0.105	0.248
	σ_3	1.000	0.643	0.074	0.467	0.900
III	σ_1	0.952	0.478	0.068	0.448	0.953
	σ_2	0.263	0.098	0.054	0.103	0.231
	σ_3	1.000	0.659	0.062	0.468	0.896
IV	σ_1	0.969	0.488	0.051	0.510	0.977
	σ_2	0.256	0.102	0.057	0.111	0.274
	σ_3	1.000	0.670	0.057	0.480	0.923

Table 3: Variable Selection for Linear Regression Models

			No. of Zeros		Fitted Models		
Design	$\sigma(X)$	MSE	C	IC	U-fit	C-fit	O-fit
			$p = 8$				
I	σ_1	0.013	4.995	0.000	0.000	0.995	0.005
	σ_2	0.123	4.970	0.005	0.005	0.965	0.030
	σ_3	0.109	4.975	0.005	0.005	0.970	0.025
II	σ_1	0.014	4.995	0.000	0.000	0.995	0.005
	σ_2	0.119	4.985	0.005	0.005	0.980	0.015
	σ_3	0.128	4.960	0.010	0.010	0.955	0.035
III	σ_1	0.013	4.965	0.000	0.000	0.980	0.020
	σ_2	0.148	4.985	0.020	0.020	0.965	0.015
	σ_3	0.096	4.985	0.000	0.000	0.985	0.015
IV	σ_1	0.012	4.990	0.000	0.000	0.990	0.010
	σ_2	0.105	4.995	0.005	0.005	0.990	0.005
	σ_3	0.093	4.970	0.000	0.000	0.980	0.020
			$p = 16$				
I	σ_1	0.017	12.975	0.000	0.000	0.975	0.025
	σ_2	0.183	12.945	0.030	0.030	0.930	0.040
	σ_3	0.113	12.985	0.010	0.010	0.975	0.015
II	σ_1	0.017	12.970	0.000	0.000	0.970	0.030
	σ_2	0.127	12.960	0.015	0.015	0.945	0.040
	σ_3	0.119	12.920	0.010	0.010	0.920	0.070
III	σ_1	0.018	12.935	0.000	0.000	0.935	0.065
	σ_2	0.152	12.945	0.015	0.015	0.940	0.045
	σ_3	0.143	12.930	0.020	0.020	0.920	0.060
IV	σ_1	0.017	12.985	0.000	0.000	0.985	0.015
	σ_2	0.120	12.960	0.005	0.005	0.955	0.040
	σ_3	0.088	12.945	0.000	0.000	0.945	0.055

Table 4: Performance of Point Estimators for Homogeneous QR Models

		$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Design		$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$	$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$	$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$
					$\varepsilon \sim N(0,1)$					
I	Bias	0.001	-0.003	-0.003	-0.006	0.000	0.001	-0.005	-0.002	0.003
	RMS	0.101	0.083	0.055	0.092	0.075	0.048	0.098	0.080	0.053
II	Bias	-0.007	0.004	-0.003	0.000	0.002	-0.005	-0.011	0.004	0.001
	RMS	0.099	0.089	0.049	0.094	0.071	0.046	0.089	0.076	0.044
III	Bias	0.006	-0.001	-0.001	-0.003	0.002	0.003	-0.003	0.000	0.002
	RMS	0.100	0.085	0.053	0.093	0.076	0.048	0.098	0.084	0.053
IV	Bias	-0.004	0.001	0.003	-0.007	0.002	0.003	-0.003	-0.003	0.000
	RMS	0.091	0.076	0.055	0.087	0.072	0.050	0.096	0.078	0.056
					$\varepsilon \sim \chi^2(3)$					
I	Bias	0.009	-0.004	0.008	0.004	0.002	0.006	-0.006	0.006	0.002
	RMS	0.145	0.125	0.074	0.195	0.168	0.100	0.284	0.242	0.147
II	Bias	0.009	-0.002	0.008	0.012	-0.001	0.002	-0.004	0.022	0.010
	RMS	0.132	0.112	0.064	0.186	0.166	0.103	0.294	0.249	0.156
III	Bias	0.004	-0.005	0.013	-0.002	0.006	0.009	-0.018	0.011	0.003
	RMS	0.148	0.124	0.075	0.197	0.168	0.101	0.288	0.247	0.149
IV	Bias	-0.002	0.003	0.014	-0.006	0.004	0.013	-0.016	0.000	0.008
	RMS	0.133	0.106	0.080	0.186	0.148	0.110	0.289	0.236	0.167
					$\varepsilon \sim t(3)$					
I	Bias	-0.002	-0.003	-0.002	-0.002	-0.001	0.001	0.003	-0.011	-0.001
	RMS	0.123	0.094	0.067	0.097	0.082	0.055	0.122	0.103	0.067
II	Bias	-0.002	0.011	0.002	-0.001	-0.003	-0.004	0.003	-0.005	-0.002
	RMS	0.117	0.103	0.079	0.096	0.077	0.053	0.123	0.095	0.062
III	Bias	-0.002	-0.002	-0.003	0.000	0.001	0.001	0.006	-0.006	-0.002
	RMS	0.123	0.096	0.068	0.096	0.084	0.056	0.120	0.106	0.068
IV	Bias	0.001	-0.002	-0.003	-0.001	0.001	0.002	0.000	-0.009	-0.001
	RMS	0.120	0.097	0.068	0.095	0.082	0.057	0.127	0.105	0.071

Table 5: Performance of Point estimators Under Heteroscedasticity

		$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
Design		$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$	$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$	$\theta_{N0}(\tau)$	$\theta_{N1}(\tau)$	$\theta_{N2}(\tau)$
					$\varepsilon \sim N(0,1)$					
I	Bias	-0.018	0.002	0.026	-0.015	0.010	0.015	-0.004	-0.001	0.023
	RMS	0.147	0.134	0.243	0.133	0.116	0.211	0.143	0.121	0.239
II	Bias	-0.016	0.012	0.006	0.008	0.005	-0.018	-0.005	0.000	0.010
	RMS	0.146	0.134	0.231	0.136	0.107	0.203	0.130	0.115	0.204
III	Bias	-0.018	0.001	0.026	-0.012	0.009	0.013	-0.003	-0.001	0.005
	RMS	0.150	0.129	0.244	0.138	0.114	0.228	0.146	0.122	0.239
IV	Bias	-0.021	0.002	0.032	-0.015	0.009	0.016	-0.002	-0.006	0.002
	RMS	0.138	0.119	0.242	0.132	0.110	0.220	0.147	0.120	0.252
					$\varepsilon \sim \chi^2(3)$					
I	Bias	0.000	-0.008	0.056	-0.015	0.001	0.054	-0.024	0.010	0.022
	RMS	0.221	0.196	0.350	0.293	0.260	0.483	0.430	0.346	0.690
II	Bias	0.003	-0.004	0.044	-0.001	-0.006	0.038	0.005	0.029	0.012
	RMS	0.201	0.181	0.296	0.299	0.252	0.503	0.447	0.350	0.759
III	Bias	0.006	-0.011	0.047	-0.009	0.006	0.050	0.000	0.007	-0.006
	RMS	0.228	0.201	0.343	0.294	0.263	0.495	0.425	0.360	0.699
IV	Bias	-0.007	-0.002	0.052	-0.016	0.007	0.058	-0.015	-0.007	0.018
	RMS	0.202	0.164	0.347	0.283	0.229	0.477	0.443	0.367	0.736
					$\varepsilon \sim t(3)$					
I	Bias	-0.021	-0.001	0.023	-0.002	0.000	0.007	0.010	-0.008	-0.003
	RMS	0.186	0.142	0.297	0.145	0.124	0.250	0.187	0.157	0.305
II	Bias	-0.015	0.007	0.021	0.008	-0.008	-0.021	0.006	-0.003	-0.002
	RMS	0.204	0.154	0.404	0.152	0.116	0.250	0.177	0.143	0.262
III	Bias	-0.023	-0.003	0.023	-0.004	0.003	0.016	0.012	-0.005	0.003
	RMS	0.186	0.149	0.304	0.148	0.128	0.242	0.184	0.165	0.308
IV	Bias	-0.015	-0.001	0.020	-0.005	0.001	0.013	0.000	-0.012	0.010
	RMS	0.181	0.149	0.299	0.144	0.126	0.243	0.188	0.163	0.304

Table 6: Size and Power of the SEL Ratio Test for Homogeneous QR Models

		$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$		
Design	b	$N(0,1)$	$\chi^2(3)$	$t(3)$	$N(0,1)$	$\chi^2(3)$	$t(3)$	$N(0,1)$	$\chi^2(3)$	$t(3)$
I	0.50	1.000	0.990	0.997	1.000	0.825	1.000	1.000	0.556	0.997
	0.75	0.833	0.575	0.754	0.884	0.346	0.870	0.866	0.194	0.740
	1.00	0.067	0.079	0.054	0.054	0.069	0.058	0.049	0.060	0.051
	1.25	0.836	0.555	0.711	0.901	0.342	0.855	0.824	0.180	0.615
	1.50	1.000	0.975	0.998	1.000	0.848	1.000	1.000	0.552	0.996
II	0.50	1.000	0.996	0.988	1.000	0.908	1.000	0.996	0.521	0.999
	0.75	0.758	0.606	0.611	0.943	0.347	0.902	0.896	0.182	0.776
	1.00	0.072	0.058	0.058	0.063	0.070	0.051	0.063	0.070	0.062
	1.25	0.854	0.597	0.736	0.933	0.359	0.871	0.893	0.205	0.710
	1.50	1.000	0.994	0.996	1.000	0.873	1.000	1.000	0.594	0.998
III	0.50	0.999	0.988	0.998	1.000	0.830	1.000	1.000	0.580	0.999
	0.75	0.830	0.573	0.726	0.865	0.318	0.855	0.839	0.190	0.701
	1.00	0.056	0.082	0.050	0.056	0.074	0.053	0.055	0.065	0.054
	1.25	0.833	0.549	0.705	0.905	0.344	0.846	0.835	0.180	0.606
	1.50	1.000	0.975	0.995	1.000	0.860	1.000	1.000	0.542	0.997
IV	0.50	1.000	0.998	0.997	1.000	0.866	1.000	1.000	0.544	0.996
	0.75	0.895	0.656	0.770	0.919	0.364	0.860	0.883	0.208	0.709
	1.00	0.059	0.060	0.054	0.056	0.057	0.049	0.057	0.071	0.060
	1.25	0.909	0.653	0.705	0.943	0.407	0.850	0.843	0.201	0.614
	1.50	1.000	0.997	0.995	1.000	0.910	1.000	1.000	0.567	0.994

Table 7: Variable Selection for the QR Model with $\tau = 0.25$

			No of Zeros		Fitted Models		
Design	Scenarios	MSE	C	IC	U-fit	C-fit	O-fit
			$p = 8$				
I	A	0.024	4.920	0.000	0.000	0.975	0.025
	B	0.027	4.860	0.000	0.000	0.960	0.040
	C	0.043	4.880	0.000	0.000	0.965	0.035
	D	0.032	5.000	0.000	0.000	1.000	0.000
	E	0.105	4.745	0.000	0.000	0.935	0.065
II	A	0.032	4.745	0.000	0.000	0.920	0.080
	B	0.032	4.870	0.000	0.000	0.960	0.040
	C	0.035	4.760	0.000	0.000	0.940	0.060
	D	0.033	4.760	0.000	0.000	0.940	0.060
	E	0.102	4.845	0.000	0.000	0.950	0.050
III	A	0.021	4.855	0.000	0.000	0.950	0.050
	B	0.029	4.620	0.000	0.000	0.895	0.105
	C	0.037	4.850	0.000	0.000	0.955	0.045
	D	0.038	4.955	0.000	0.000	0.990	0.010
	E	0.105	4.815	0.000	0.000	0.955	0.045
IV	A	0.019	4.855	0.000	0.000	0.955	0.045
	B	0.023	4.785	0.000	0.000	0.935	0.065
	C	0.038	4.940	0.000	0.000	0.980	0.020
	D	0.029	4.965	0.000	0.000	0.985	0.015
	E	0.100	4.795	0.000	0.000	0.930	0.070
			$p = 16$				
I	A	0.047	12.030	0.000	0.000	0.855	0.145
	B	0.126	12.070	0.020	0.015	0.815	0.170
	C	0.176	12.505	0.055	0.045	0.875	0.080
	D	0.118	12.500	0.030	0.030	0.880	0.090
	E	0.156	12.055	0.010	0.010	0.830	0.160
II	A	0.061	11.945	0.005	0.005	0.800	0.195
	B	0.064	12.110	0.010	0.010	0.835	0.155
	C	0.428	12.300	0.120	0.090	0.795	0.115
	D	0.130	12.310	0.035	0.035	0.850	0.115
	E	0.210	11.240	0.020	0.015	0.735	0.250
III	A	0.043	12.015	0.000	0.000	0.850	0.150
	B	0.212	12.060	0.035	0.020	0.825	0.155
	C	0.201	12.500	0.070	0.065	0.835	0.100
	D	0.066	12.440	0.010	0.010	0.890	0.100
	E	0.233	11.990	0.040	0.030	0.785	0.185
IV	A	0.039	12.080	0.000	0.000	0.870	0.130
	B	0.043	12.210	0.000	0.000	0.875	0.125
	C	0.147	12.465	0.055	0.055	0.860	0.085
	D	0.099	12.310	0.020	0.015	0.865	0.120
	E	0.181	11.890	0.010	0.010	0.800	0.190

Table 8: Variable Selection for the QR Model with $\tau = 0.50$

			No of Zeros		Fitted Models		
Design	Scenarios	MSE	C	IC	U-fit	C-fit	O-fit
			$p = 8$				
I	A	0.020	4.800	0.000	0.000	0.945	0.055
	B	0.025	4.820	0.000	0.000	0.935	0.065
	C	0.084	4.970	0.005	0.005	0.985	0.010
	D	0.026	4.920	0.000	0.000	0.975	0.025
	E	0.008	4.745	0.000	0.000	0.920	0.080
II	A	0.023	4.755	0.000	0.000	0.935	0.065
	B	0.022	4.700	0.000	0.000	0.920	0.080
	C	0.111	4.975	0.020	0.020	0.975	0.005
	D	0.024	4.830	0.000	0.000	0.950	0.050
	E	0.009	4.310	0.000	0.000	0.760	0.240
III	A	0.019	4.810	0.000	0.000	0.940	0.060
	B	0.024	4.795	0.000	0.000	0.940	0.060
	C	0.083	4.920	0.005	0.005	0.975	0.020
	D	0.024	4.945	0.000	0.000	0.985	0.015
	E	0.009	4.895	0.000	0.000	0.960	0.040
IV	A	0.016	4.925	0.000	0.000	0.980	0.020
	B	0.019	4.850	0.000	0.000	0.955	0.045
	C	0.080	4.980	0.010	0.010	0.985	0.005
	D	0.019	4.940	0.000	0.000	0.980	0.020
	E	0.006	4.915	0.000	0.000	0.975	0.025
			$p = 16$				
I	A	0.066	12.460	0.020	0.020	0.885	0.095
	B	0.053	12.280	0.010	0.010	0.835	0.155
	C	0.435	13.000	0.100	0.075	0.925	0.000
	D	0.161	12.075	0.045	0.035	0.835	0.130
	E	0.047	11.300	0.010	0.010	0.695	0.295
II	A	0.040	12.085	0.000	0.000	0.835	0.165
	B	0.199	12.150	0.025	0.020	0.840	0.140
	C	0.449	12.905	0.125	0.095	0.885	0.020
	D	0.074	12.325	0.020	0.020	0.835	0.145
	E	0.058	11.020	0.020	0.020	0.670	0.310
III	A	0.037	12.150	0.000	0.000	0.865	0.135
	B	0.046	12.060	0.005	0.005	0.845	0.150
	C	0.287	12.925	0.075	0.055	0.925	0.020
	D	0.128	12.585	0.040	0.035	0.880	0.085
	E	0.069	11.780	0.015	0.010	0.775	0.215
IV	A	0.046	12.115	0.005	0.005	0.830	0.165
	B	0.061	11.685	0.005	0.005	0.780	0.215
	C	0.156	12.820	0.035	0.035	0.930	0.035
	D	0.117	11.985	0.035	0.030	0.800	0.170
	E	0.060	11.350	0.015	0.015	0.685	0.300

Table 9: Variable Selection for the QR Model with $\tau = 0.75$

			No of Zeros		Fitted Models		
Design	Scenarios	MSE	C	IC	U-fit	C-fit	O-fit
			$p = 8$				
I	A	0.023	4.890	0.000	0.000	0.970	0.030
	B	0.026	4.780	0.000	0.000	0.935	0.065
	C	0.238	4.975	0.045	0.045	0.950	0.005
	D	0.033	4.810	0.000	0.000	0.945	0.055
	E	0.071	4.860	0.000	0.000	0.960	0.040
II	A	0.022	4.720	0.000	0.000	0.915	0.085
	B	0.037	4.905	0.005	0.005	0.955	0.040
	C	0.190	5.000	0.030	0.030	0.970	0.000
	D	0.024	4.830	0.000	0.000	0.950	0.050
	E	0.066	4.700	0.000	0.000	0.905	0.095
III	A	0.022	4.970	0.000	0.000	0.990	0.010
	B	0.025	4.850	0.000	0.000	0.950	0.050
	C	0.176	5.000	0.010	0.010	0.990	0.000
	D	0.032	4.940	0.000	0.000	0.985	0.015
	E	0.067	4.875	0.000	0.000	0.955	0.045
IV	A	0.019	4.870	0.000	0.000	0.960	0.040
	B	0.024	4.870	0.000	0.000	0.960	0.040
	C	0.217	4.990	0.040	0.040	0.955	0.005
	D	0.030	4.890	0.000	0.000	0.975	0.025
	E	0.075	4.805	0.000	0.000	0.925	0.075
			$p = 16$				
I	A	0.055	11.945	0.005	0.005	0.850	0.145
	B	0.173	11.705	0.045	0.035	0.775	0.190
	C	0.392	12.985	0.100	0.085	0.910	0.005
	D	0.102	12.075	0.015	0.015	0.845	0.140
	E	0.093	12.125	0.010	0.010	0.795	0.195
II	A	0.063	11.445	0.000	0.000	0.760	0.240
	B	0.183	12.090	0.030	0.020	0.810	0.170
	C	0.718	12.895	0.190	0.160	0.810	0.030
	D	0.193	12.225	0.040	0.035	0.820	0.145
	E	0.095	11.700	0.010	0.010	0.780	0.210
III	A	0.047	12.260	0.000	0.000	0.880	0.120
	B	0.280	11.955	0.045	0.030	0.790	0.180
	C	0.430	12.940	0.125	0.120	0.870	0.010
	D	0.168	11.985	0.055	0.055	0.795	0.150
	E	0.097	12.190	0.010	0.010	0.830	0.160
IV	A	0.030	12.475	0.000	0.000	0.925	0.075
	B	0.115	11.990	0.020	0.010	0.835	0.155
	C	0.436	12.990	0.135	0.125	0.870	0.005
	D	0.103	11.900	0.015	0.015	0.815	0.170
	E	0.088	12.040	0.005	0.005	0.805	0.190