SEMIPARAMETRIC ESTIMATION OF NON-IGNORABLE MISSINGNESS WITH REFRESHMENT SAMPLE

Jianfei Zheng, Jing Wang, Lan Xue and Annie Qu

BeiGene, University of Illinois at Chicago, Oregon State University, and University of California, Irvine

Supplementary Material

This supplement material contains an extension of the proposed method to the case where a binary covariate is included. In addition, results from additional simulation studies, necessary lemmas, and detailed proofs of the main theorems are also presented.

S1 Binary Covariates

In the main paper, the missing mechanism is assumed to depend only on the response variables through an additive non-ignorable model. Oftentimes, this dependency involves covariates. Therefore it is desirable to incorporate covariates into our missing mechanism model. In this section, we extend our semiparametric model to incorporate one time-invariant binary covariate. This covariate can be related to the responses and the missing mechanism.

Let Y_{i1} and Y_{i2} denote the ith responses at the first and second waves respectively. Let X_i represent a time-invariant categorical covariate for the ith observation. For simplicity, we assume that X_i only takes value in two levels, 0 and 1. Let W_i be an indicator variable, with $W_i = 0$ indicating that Y_{i2} is missing. An additive missing model is assumed for W_i,

$$P(W_i = 1 \mid y_{i1}, y_{i2}, x_i) = \text{logistic}(\beta_0 + \beta_1 y_{i1} + \beta_2 y_{i2} + \beta_3 x_i).$$

This logistic attrition model allows for a straightforward interpretation of the covariate as the main effect on the odds ratio being observed. A more complex attrition model can be specified according to Hirano et al. (2001) as

$$P(W = 1 \mid y_1, y_2, x) = g(\kappa_0(x) + \kappa_1(y_1, x) + \kappa_2(y_2, x)),$$

where g is a monotone function taking on values in the interval $(0, 1)$, and $\kappa_1(\cdot)$, $\kappa_2(\cdot)$, $\kappa_3(\cdot)$ are arbitrary functions of the responses and the covariate. It is important, however, to note that no interaction terms between y_1 and y_2 are allowed in this additive model.

In addition, a refreshment sample is also included in the second wave. Table 1 shows the observed data in this scenario. Similar to the no-covariate case, the observed panel data can be separated into two sets according to the values of

W. The complete set consists of observations with $W = 1$, and we fully observe every variable in this set. The rest of the panel data then form the incomplete set. Again, the goal is to understand the attrition process by estimating the attrition parameters from the data that we observe in Table 1.

S1.1 Method

Estimation of the attrition parameters can be obtained through Hirano et al. (2001)'s two constraints on the covariate x. In particular, we have

$$\begin{aligned}
\int \frac{P(W=1 \mid x)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 x)} f(y_1, y_2 \mid W=1, x) dy_2 &= f_1(y_1 \mid x), \\
\int \frac{P(W=1 \mid x)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 x)} f(y_1, y_2 \mid W=1, x) dy_1 &= f_2(y_2 \mid x).
\end{aligned}$$

The idea for estimating the $\beta = (\beta_{\mathbf{0}}, \cdots, \beta_{\mathbf{3}})$ in this scenario is similar to what we have done in the no-covariate case. We can consider the previous no-covariate situation as a special case where the covariate X has only one level. With one binary covariate, we can separate the data into two subsets defined by the levels of X. In each subset, we construct two constraints as follows. For $X = 0$, we have

$$\begin{aligned}
\int \frac{P(W=1 \mid X=0)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1, X=0) dy_2 &= f_1(y_1 \mid X=0), \\
\int \frac{P(W=1 \mid X=0)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2)} f(y_1, y_2 \mid W=1, X=0) dy_1 &= f_2(y_2 \mid X=0).
\end{aligned}$$

In addition for $X = 1$,

$$\int \frac{P(W=1 \mid X=1)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3)} f(y_1, y_2 \mid W=1, X=1) dy_2 = f_1(y_1 \mid X=1),$$
$$\int \frac{P(W=1 \mid X=1)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3)} f(y_1, y_2 \mid W=1, X=1) dy_1 = f_2(y_2 \mid X=1).$$

The true attrition parameters β^0 are the only set of parameters that satisfy the above constraints. As a result, the estimates for these parameters can be obtained by minimizing the distance between the conditional density functions on both sides of these four constraints. The estimation procedure starts with estimating the conditional density components in the above constraints. In each subset ($i = 0$ or 1), we estimate

$$f(y_1, y_2 \mid W=1, X=i), \qquad P(W=1 \mid X=i),$$
$$f_1(y_1 \mid X=i), \qquad f_2(y_2 \mid X=i).$$

We consider the estimation of these density components using the kernel density method. We consider the estimation of these quantities for $X = 1$, and similar estimators can be constructed for $X = 0$. First, we estimate the conditional joint distribution $f(y_1, y_2 \mid W=1, X=1)$ as

$$\widehat{f}_H(y_1, y_2 \mid W=1, X=1) = \widehat{f}_H(\mathbf{y} \mid W=1, X=1) = \frac{1}{n_{11}} \sum_{i=1}^{n_{11}} K_H(\mathbf{y} - \mathbf{Y}_i),$$

where $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$, $i = 1, 2, ..., n_{11}$ indexes the data points with both $W = 1$ and $X = 1$, and H is a 2×2 bandwidth matrix that is symmetric and positive definite. Additionally, $K_H(\mathbf{y}) = |H|^{-1/2} K(H^{-1/2}\mathbf{y})$, where K is the bivariate normal kernel function defined as $K(\mathbf{y}) = (2\pi)^{-1} exp(-\mathbf{y}^T\mathbf{y}/2)$. Next, $P(W = 1 \mid X = 1)$ can be consistently estimated by $\widehat{P}(W = 1 \mid X = 1) = n_{11}/N_1$, where N_1 is the number of observations with $X = 1$. For a given β, the estimator of the conditional joint density $f(y_1, y_2 \mid X = 1)$ is given as

$$\widetilde{f}(y_1, y_2 \mid X = 1, \beta) = \frac{\widehat{P}(W = 1 \mid X = 1)}{\text{logistic}(\beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3)} \widehat{f}_H(y_1, y_2 \mid W = 1, X = 1).$$

The conditional density of Y_1 given $X = 1$ can be computed by integrating the conditional joint distribution $\widetilde{f}(y_1, y_2 \mid X = 1, \beta)$ with respect to y_2. This can be numerically approximated as

$$\begin{aligned}
\widetilde{f}_1(y_1 \mid X = 1, \beta) &= \int \widetilde{f}(y_1, y_2 \mid X = 1, \beta) dy_2 \\
&\approx \sum_{i=1}^{n_{grid}} \widetilde{f}(y_1, y_{2i} \mid X = 1, \beta) \times \Delta y_2 \\
&= \sum_{i=1}^{n_{grid}} \widetilde{f}(y_1, y_{2i} \mid \beta) \times \frac{range(y_2)}{n_{grid}},
\end{aligned}$$

where y_{2i} is the ith grid point on Y_2 and n_{grid} denotes the number of grid points in the 2-dimensional kernel density estimator. Similarly, for a given y_2, the conditional density $\widetilde{f}_2(y_2 \mid X = 1, \beta)$ can be defined in the same manner. The conditional density estimates $\widetilde{f}_1(y_1 \mid X = 1, \beta)$ and $\widetilde{f}_2(y_2 \mid X = 1, \beta)$ are semi-parametric estimators that rely on the attrition model. They consistently estimate the true marginal densities only when the attrition model is correctly specified.

Let $\{y_{i1}\}_{i=1}^{N_1}$ be the first wave responses with $X = 1$ and $\{y_{i2}^r\}_{i=1}^{n_1}$ be the refreshment sample with $X = 1$. We define the following one-dimensional kernel density estimators:

$$\widehat{f}_1(y_1 \mid X = 1) = \frac{1}{N_1} \sum_{i=1}^{N_1} K_{h_1}(y_1 - y_{i1}), \qquad \widehat{f}_2(y_2 \mid X = 1) = \frac{1}{n_1} \sum_{i=1}^{n_1} K_{h_2}(y_2 - y_{i2}^r),$$

where K is the univariate normal density function and $K_{h_i}(y) = h_i^{-1} K(y/h_i)$, with h_i being the corresponding bandwidth for $i = 1, 2$.

In the subset with $X = 0$, a set of similar conditional density estimators can be constructed, and they are denoted as

$$\widetilde{f}_1(y_1 \mid X = 0, \beta), \qquad \widetilde{f}_2(y_2 \mid X = 0, \beta),$$
$$\widehat{f}_1(y_1 \mid X = 0), \qquad \widehat{f}_2(y_2 \mid X = 0).$$

The objective function $M(\beta)$ takes the form of the mean squared differences between the corresponding conditional density functions from the left and righthand sides of the four constraints,

$$
\begin{aligned}
M(\beta) = &\quad M_{N_0}(\beta) + M_{n_0}(\beta) + M_{N_1}(\beta) + M_{n_1}(\beta) \\
= &\quad \tfrac{1}{N_0} \textstyle\sum_{i=1}^{N_0} \left[\widetilde{f}_1(y_{i1} \mid X = 0, \beta) - \widehat{f}_1(y_{i1} \mid X = 0)\right]^2 \\
&+ \tfrac{1}{n_0} \textstyle\sum_{i=1}^{n_0} \left[\widetilde{f}_2(y^r_{i2} \mid X = 0, \beta) - \widehat{f}_2(y^r_{i2} \mid X = 0)\right]^2 \\
&+ \tfrac{1}{N_1} \textstyle\sum_{i=1}^{N_1} \left[\widetilde{f}_1(y_{i1} \mid X = 1, \beta) - \widehat{f}_1(y_{i1} \mid X = 1)\right]^2 \\
&+ \tfrac{1}{n_1} \textstyle\sum_{i=1}^{n_1} \left[\widetilde{f}_2(y^r_{i2} \mid X = 1, \beta) - \widehat{f}_2(y^r_{i2} \mid X = 1)\right]^2.
\end{aligned}
$$

Notice that there are four comparisons since there are four constraints. The vector of semi-parametric estimators of the attrition parameters is the minimizer of the objective function as $\widehat{\beta} = \arg\min_{\beta} M(\beta)$.

S1.2 Simulation Results

We use simulation to demonstrate the finite sample performance of the semi-parametric estimators in this one-covariate case. We generate data from the following model

$$
\begin{aligned}
Y_{i1} &= 2 + X_i + \epsilon_{i1}, \\
Y_{i2} &= 1 + 3X_i + \epsilon_{i2}, \\
\underline{\epsilon} &= \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix} \right).
\end{aligned} \tag{S1.1}
$$

The additive missing model is set up as

$$
P(W_i = 1 \mid Y_{i1}, Y_{i2}, X_i) = \text{logistic}(-0.8 + 0.2Y_{i1} + 0.4Y_{i2} - 1.4X_i). \tag{S1.2}
$$

The attrition parameters are set up such that given any value of the covariate X, there is a non-zero probability to observe the responses Y_1 and Y_2 almost everywhere on their corresponding support. That is, given $X = x$, the support of $f(y_1, y_2 \mid W = 1, x)$ coincides with the support of $f(y_1, y_2 \mid x)$. In this particular additive missing model setting, we are able to control the probability of Y_2 being missing so that it is about 50% on average and ranges from 20% to 80% given either level of X.

The attrition parameters are estimated by the semi-parametric method for each sample. The squared bias and variance are calculated for each estimator based on those 1000 estimates, and the corresponding MSE is computed. Figure

1 shows plots of the MSE versus panel and refreshment sample sizes. The X-axis represents the combination of panel size and refreshment sample size. The dashed, dot-dash, and solid lines represent the squared bias, variance, and MSE respectively. The decreasing trends of the empirical MSE suggest that the semi-parametric estimators are consistent.

S2 Additional simulation

In this section, we present additional simulation studies to investigate the effects of different model parameters such as marginal variances and missing mechanism parameters, on the performance of the semi-parametric estimation method.

S2.1 Effect of marginal variances

To investigate the effect of marginal variances on the variability of the semi-parametric estimator, we generate data from a bivariate Normal distribution with marginal means of $\mathbf{0}$ and correlation coefficient of 0.5, but with marginal variances at three different levels ($\sigma_1^2, \sigma_2^2 = 1, 5$ or 10). Table 2 reports the variances of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ computed through both asymptotic formula and empirical variance using simulated data. It shows that the variability of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ decreases as the marginal variance of first wave σ_1^2 or the second wave σ_2^2 increases.

In addition, the variances calculated using the asymptotic formula agree with the empirical variance, which validates the asymptotic variance given in Theorem 3 of the paper. However, the asymptotic formula gives slightly larger standard errors than the empirical SEs obtained by simulation. This may be due to the fact that all higher-order terms are ignored and a Taylor expansion is repeatedly used to simplify integrals involving kernel densities in the development of the asymptotic theory.

Similar simulations were also conducted to investigate the effects of marginal means and the correlation coefficient on the estimation performance. The simulation result is not reported here. But it shows that the value of marginal means plays an ignorable effect on the variability for both estimators. However, the variability increase as the correlation coefficient increases across different marginal variance combinations. In conclusion, less correlation and more variability in the marginal distributions result in more stable estimates of attrition parameters.

S2.2 Effect of The Rotation in Missing Direction

For the logistic attrition model, the missingness probability is constant along any line where $\beta_0 + \beta_1 y_1 + \beta_2 y_2$ is a fixed constant. In particular, the line $\beta_0 + \beta_1 y_1 + \beta_2 y_2 = 0$ is called the reference line, which corresponds to 50% missingness of Y_2. The vector $(\beta_1, \beta_2)^T$ gives the direction that is perpendicular to the reference line, and θ denotes the rotation angle of vector $(\beta_1, \beta_2)^T$ as shown in Figure 2. We refer to this vector as the perpendicular vector (PV). The missing probability of Y_2 decreases in the direction of PV with the rate of decreasing depending on the length of PV calculated as $\sqrt{\beta_1^2 + \beta_2^2}$.

Figure 3 considers two different lengths of PVs, 0.5 and 1. A length of 0.5 gives a gradual missingness pattern and the probability of being observed is away from 0 and 1. A length of 1 shows a dramatic missing pattern where part of the data is almost always observed and the other part is missing most of the time. The X-axis represents different values for rotation angle θ. There are 8 equally spaced rotations with angles ranging from $\frac{3}{2}\pi$ to $\frac{5}{4}\pi$ counterclockwise. in Figure 3, the angles are denoted as the rotation index $1, \cdots, 8$ respectively.

Figure 3 shows the semi-parametric estimator has a larger variance when data are dramatically missing (length of PV = 1). Furthermore, it reveals that minimum asymptotic standard errors are achieved when the PV parallels the major axis of the population contour. In our bivariate Normal case, the joint distribution has a positive correlation coefficient and marginal variances are equal, which results in a population contour with a 45-degree major axis. Two rotation scenarios give the parallel relationship between the PV and the 45-degree major axis, namely scenarios with rotation index of 4 and 8 (i.e. $\frac{1}{4}\pi$ and $\frac{5}{4}\pi$ respectively). When the PV is perpendicular to the major axis, asymptotic standard errors reach the maximum (i.e. $\frac{7}{4}\pi$ and $\frac{3}{4}\pi$ for rotation index 2 and 6).

S3 Technical Lemmas and Proofs

Let $\widetilde{f}_1(y_1 \mid \beta)$ and $\widehat{f}_1(y_1)$ be the semi-parametric and non-parametric estimators of the marginal density of the first wave $f_1(y_1)$ as in the objective function $M_{N,n}(\beta)$. In addition, let

$$f_1(y_1 \mid \beta) = \int \frac{f\left(y_1, y_2 \mid W=1\right) P\left(W=1\right)}{1/\left(1+\exp\left(-\beta_0-\beta_1 y_1-\beta_2 y_2\right)\right)} dy_2, \tag{S3.3}$$

which is defined similarly as $\widetilde{f}_1(y_1 \mid \beta)$ but with true quantities for $f\left(y_1, y_2 \mid W=1\right)$ and $P\left(W=1\right)$ instead of the estimated ones. Let

$$\begin{aligned} A_{1\beta}(y_1) &= f_1(y_1 \mid \beta) - f_1(y_1), \quad B_{1\beta}(y_1) = \widetilde{f}_1(y_1 \mid \beta) - f_1(y_1 \mid \beta), \\ C_1(y_1) &= f_1(y_1) - \widehat{f}_1(y_1). \end{aligned}$$

In the same manner, for the second wave, we define

$$\begin{aligned} A_{2\beta}(y_2) &= f_2(y_2 \mid \beta) - f_2(y_2), \quad B_{2\beta}(y_2) = \widetilde{f}_2(y_2 \mid \beta) - f_2(y_2 \mid \beta), \\ C_2(y_2) &= f_2(y_2) - \widehat{f}_2(y_2), \end{aligned}$$

where $f_2(y_2 \mid \beta)$ is defined similarly as $f_1(y_1 \mid \beta)$ in (S3.3), but integrating y_1

out instead.

Lemma 1. *For any θ in a compact set Θ, let $x \mapsto f_\theta(x)$ be a given measurable function. Suppose $\theta \mapsto f_\theta(x)$ is continuous for every x and suppose that there exists a function F such that $|f_\theta| \leq F$ for every $\theta \in \Theta$, and $PF < +\infty$, then $\sup_{\theta\in\Theta} |P_n f_\theta - P f_\theta| \xrightarrow{P} 0$.*

This result is shown in section 19.2 of Van der Vaart (2000).

Lemma 2. *Under assumptions (A1)-(A3), one has*

$$\sup_{\beta\in\Theta} \left| P_N A_{1\beta}^2 - P A_{1\beta}^2 \right| \xrightarrow{P} 0.$$

Proof of Lemma 2. Under assumptions (A1)-(A3), for every β, one has

$$\begin{aligned} A_{1\beta}^2(y_1) &= \left[\int f(y_1, y_2 \mid W = 1) P(W = 1)(1 + \exp(-\beta \mathbf{y})) dy_2 - f_1(y_1)\right]^2 \\ &\leq F(y_1), \end{aligned}$$

for some $F(y_1)$ that only depends on y_1 and $E[F(Y_1)] < +\infty$. The result follows from Lemma 1. □

Lemma 3. *Under conditions (A1)-(A6), one has*

$$\sup_{\beta\in\Theta} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[B_{1\beta}(Y_{i1}) + B_{1\beta}^2(Y_{i1}) + \frac{\partial}{\partial\beta} B_{1\beta}(Y_{i1}) + C_1(Y_{i1}) + C_1^2(Y_{i1}) \right] \right\} = o_p(1).$$

Proof of Lemma 3. One notes that, for any $\beta \in \Theta$,

$$\begin{aligned} B_{1\beta}(y_1) &= \widetilde{f}(y_1 \mid \beta) - f(y_1 \mid \beta) = \int \frac{\widehat{P}\widehat{f}_H - Pf}{\text{logistic}(\beta \mathbf{y})} dy_2 \\ &= \int \frac{(\widehat{P} - P)\widehat{f}_H + P(\widehat{f}_H - f)}{\text{logistic}(\beta \mathbf{y})} dy_2, \end{aligned}$$

where $\widehat{P} - P = \frac{1}{N}\sum_{i=1}^N I(W_i = 1) - P(W = 1) = o_p(1)$ by the weak law of large numbers. In addition, the uniform convergence of the bivariate kernel density estimator in Devroye and Wagner (1980) gives that

$$\sup_{y_1, y_2} \left| \widehat{f}_H(y_1, y_2 \mid W = 1) - f(y_1, y_2 \mid W = 1) \right| \xrightarrow{P} 0 \quad \text{as} \quad n_c \to \infty. \quad \text{(S3.4)}$$

Therefore, $\sup_{y_1} |B_{1\beta}(y_1)| = o_p(1)$. Similarly, $\sup_{y_1} |\frac{\partial}{\partial \beta} B_{1\beta}(y_1)| = o_p(1)$. Thus assumptions (A1)-(A3) give that

$$\sup_{\beta \in \Theta} \frac{1}{N} \sum_{i=1}^N \left\{ B_{1\beta}(Y_{i1}) + B_{1\beta}^2(Y_{i1}) + \frac{\partial}{\partial \beta} B_{1\beta}(Y_{i1}) \right\} = o_p(1).$$

By the uniform convergence of the univariate density estimator given in Theorem A of Silverman et al. (1978), we have $\sup_{y_1} |C_1(y_1)| = \sup_{y_1} \left| \widehat{f}_1(y_1) - f_1(y_1) \right| \xrightarrow{\text{a.s.}} 0 \quad \text{as} \quad N \to \infty$. As a result, $\frac{1}{N}\sum_{i=1}^N C_1(y_{i1}) \le \sup_{y_1} |C_1(y_1)| = o_p(1)$ and $\frac{1}{N}\sum_{i=1}^N C_1^2(y_{i1}) \le \sup_{y_1} |C_1(y_1)|^2 = o_p(1)$. □

Lemma 4. *Under (A1) – (A6), one has*

$$\sup_{\beta\in\Theta}\left|M_N(\beta) - E\left[f_1(Y_1 \mid \beta) - f_1(Y_1)\right]^2\right| \xrightarrow{P} 0.$$

Proof of Lemma 4. Note that,

$$\begin{aligned} M_N(\beta) &= \frac{1}{N}\sum_{i=1}^{N}\left[A_{1\beta}(Y_{i1}) + B_{1\beta}(Y_{i1}) + C_1(Y_{i1})\right]^2 = \frac{1}{N}\sum_{i=1}^{N} A_{1\beta}^2(Y_{i1}) \\ &+ \frac{1}{N}\sum_{i=1}^{N}\left[2A_{1\beta}(Y_{i1})B_{1\beta}(Y_{i1}) + A_{\beta}(Y_{i1})C_1(Y_{i1}) + B_{1\beta}(Y_{i1})C_1(Y_{i1}) + B_{1\beta}^2(Y_{i1}) + C_1^2(Y_{i1})\right]. \end{aligned}$$

Then the proof follows from Lemmas 2 and 3. □

Lemma 5. *Under (A1) – (A6), one has*

$$\sup_{\beta\in\Theta}\left|M_{N,n}(\beta) - E\left[f_1(Y_1 \mid \beta) - f_1(Y_1)\right]^2 - E\left[f_2(Y_2 \mid \beta) - f_2(Y_2)\right]^2\right| \xrightarrow{P} 0.$$

Proof of Lemma 5. The proof follows similarly to the proof of Lemma 4. □

S3.1 Proof of Theorem 2

By Lemma 1 of the paper, for almost all $(y_1, y_2) \in S$, $\beta = \beta^0$ is the unique set of parameters that satisfy $f_1(y_1 \mid \beta^0) - f_1(y_1) = 0$ and $f_2(y_2 \mid \beta^0) - f_2(y_2) = 0$. Thus β_0 is the unique minimizer of $E\left[f_1(Y_1 \mid \beta) - f_1(Y_1)\right]^2 + E\left[f_2(Y_2 \mid \beta) - f_2(Y_2)\right]^2$. Combining with Lemma 5, the consistency of $\widehat{\beta}$ follows from Theorem 5.7 of Van der Vaart (2000).

S3.2 Proof of Theorem 3

The asymptotic properties of $\widehat{\beta}$ can be evaluated through the form of a Z-estimator by taking the derivative of $M_{N,n}(\beta)$. There are two parts in $M_{N,n}(\beta)$, namely $M_N(\beta)$ and $M_n(\beta)$. In the following, we will tackle each part separately and put them back together at the end to obtain the asymptotic Normality of $\widehat{\beta}$.

The First Part. Using the notation in Equation (S3.4), the first part of $M_{N,n}(\beta)$ can be decomposed as,

$$M_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{e_1(Y_{i1})\left[A_{1\beta}(Y_{i1}) + B_{1\beta}(Y_{i1}) + C_1(Y_{i1})\right]\right\}^2.$$

Then the first order derivative of $M_N(\beta)$ is

$$\begin{aligned}\varphi_N(\beta) &= \frac{\partial M_N(\beta)}{\partial\beta}\\ &= \frac{2}{N}\sum_{i=1}^{N} e_1^2(Y_{i1})\left[A_{1\beta}(Y_{i1}) + B_{1\beta}(Y_{i1}) + C_1(Y_{i1})\right]\frac{\partial}{\partial\beta}\left[A_{1\beta}(Y_{i1}) + B_{1\beta}(Y_{i1})\right].\end{aligned}$$

When $\varphi_N(\beta)$ is evaluated at the truth β^0, $f_1(Y_{i1} \mid \beta^0) = f_1(Y_{i1})$, and $A_{1\beta^0}(Y_{i1}) = 0$. Then Lemma 3 entails that

$$\begin{aligned}\varphi_N\left(\beta^0\right) = & \frac{2}{N}\sum_{i=1}^N \left[e_1^2\left(Y_{i1}\right)\left\{\frac{\partial}{\partial\beta}A_{1\beta^0}(Y_{i1}) + \frac{\partial}{\partial\beta}B_{1\beta^0}(Y_{i1})\right\}\left\{B_{1\beta_0}(Y_{i1}) + C_1(Y_{i1})\right\}\right] \\ = & \frac{1}{N}\sum_{i=1}^N \left[2e_1^2\left(Y_{i1}\right)\frac{\partial}{\partial\beta}A_{1\beta^0}(Y_{i1})\left(B_{1\beta^0}(Y_{i1}) + C_1(Y_{i1})\right)\right](1 + o_p(1)).\end{aligned}$$

Here $\frac{\partial}{\partial\beta}A_{1\beta^0}(Y_{i1}) = g\left(Y_{i1}\right) = \left[g_1\left(Y_{i1}\right), g_2\left(Y_{i1}\right), g_3\left(Y_{i1}\right)\right]^T$ with

$$\begin{aligned}g_1\left(Y_{i1}\right) &= -\int f\left(Y_{i1}, y_2 \mid W = 1\right) P\left(W = 1\right)\exp\left(-\beta_0^0 - \beta_1^0 Y_{i1} - \beta_2^0 y_2\right) dy_2, \\ g_2\left(Y_{i1}\right) &= -\int Y_{i1} f\left(Y_{i1}, y_2\right)\left\{1 + \exp\left(\beta_0^0 + \beta_1^0 Y_{i1} + \beta_2^0 y_2\right)\right\}^{-1} dy_2, \\ g_3\left(Y_{i1}\right) &= -\int y_2 f\left(Y_{i1}, y_2\right)\left\{1 + \exp\left(\beta_0^0 + \beta_1^0 Y_{i1} + \beta_2^0 y_2\right)\right\}^{-1} dy_2.\end{aligned}$$

Define a function

$$T_1\left(x, y, z, w\right) = \int \frac{wK_{h_1}\left(z - x\right)K_{h_2}\left(y_2 - y\right)}{1/\left(1 + \exp\left(-\beta_0^0 - \beta_1^0 z - \beta_2^0 y_2\right)\right)} dy_2 - K_{h_1}\left(z - x\right).$$

Then we have $\varphi_N\left(\beta^0\right) = \frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \left[2e_1^2\left(Y_{i1}\right) g\left(Y_{i1}\right) T_1\left(y_{j1}, y_{j2}, Y_{i1}, w_j\right)\right] + o_p(1)$.

Let $\mathbf{X}_i = \left[Y_{i1}, Y_{i2}, W_i\right]^T$ and $\mathbf{X}_j = \left[Y_{j1}, Y_{j2}, W_j\right]^T$ be independent samples from the panel. Let $h\left(\mathbf{X}_i, \mathbf{X}_j\right) = \tilde{h}(\mathbf{X}_i, \mathbf{X}_j) + \tilde{h}(\mathbf{X}_j, \mathbf{X}_i)$, where $\tilde{h}(\mathbf{X}_i, \mathbf{X}_j) =$

$e_1^2(Y_{i1})\,g(Y_{i1})\,T_1(Y_{j1},Y_{j2},Y_{i1},w_j)$. Then h is a symmetric function and

$$\varphi_N\left(\beta^0\right) \approx \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} h\left(\mathbf{X}_i,\mathbf{X}_j\right)$$

is a V-statistic.

Lemma 6. *Let* $h_1(\mathbf{X}_i) = E\left[h(\mathbf{X}_i,\mathbf{X}_j) \mid \mathbf{X}_i\right]$ *and* $\Sigma_1 = Var\left[h_1(\mathbf{X})\right]$. *Then under assumptions (A1)-(A6), one has* $\sqrt{N}\varphi_N(\beta^0) \sim N(\mathbf{0}, 4\Sigma_1)$.

Proof of Lemma 6. Note that

$$E\left[h\left(\mathbf{X}_i,\mathbf{X}_j\right)\right] = E\left[E\left(\tilde{h}\left(\mathbf{X}_i,\mathbf{X}_j\right)|\mathbf{X}_i\right)\right] + E\left[E\left(\tilde{h}\left(\mathbf{X}_j,\mathbf{X}_i\right)|\mathbf{X}_i\right)\right] = I + II.$$

For I, conditional on $\mathbf{X}_i$,

$$\begin{aligned} I &= E\left\{E\left[e_1^2(Y_{i1})\,g(Y_{i1})\,T_1(Y_{j1},Y_{j2},Y_{i1},W_j) \mid \mathbf{X}_i\right]\right\} \\ &= E\left\{e_1^2(Y_{i1})\,g(Y_{i1})\int T_1(y_{j1},y_{j2},Y_{i1},w_j)\,f(\mathbf{X}_j \mid \mathbf{X}_i)\,d\mathbf{X}_j\right\}. \end{aligned}$$

Let $u_1 = \frac{y_{j1}-Y_{i1}}{h_1}$ and $u_2 = \frac{y_{j2}-y_2}{h_2}$. With the change of variable and the Taylor expansion, one has

$$\begin{aligned} I \approx & E\left\{e_1^2(Y_{i1})\,g(Y_{i1})\left(\int\int\int K(u_1)\,K(u_2)\,du_1du_2 f(Y_{i1},y_2)\,dy_2 - f_1(Y_{i1})\right)\right\} \\ = & E\left\{e_1^2(Y_{i1})\,g(Y_{i1})\left(\int f(Y_{i1},y_2)\,dy_2 - f_1(Y_{i1})\right)\right\} = 0. \end{aligned}$$

Similarly, $II = o(1)$. Thus $E\left[h\left(\mathbf{X}_i, \mathbf{X}_j\right)\right] = o(1)$. For the variance, similar calculations give that

$$\begin{aligned}
h_1\left(\mathbf{X}_i\right) &= E\left[e_1^2\left(Y_{j1}\right) g\left(Y_{j1}\right) T_1\left(Y_{i1}, Y_{i2}, Y_{j1}, Wi\right) \mid \mathbf{X}_i\right] \\
&= \int\int\int e_1^2\left(Y_{i1}\right) g\left(Y_{i1}\right) \frac{W_i K\left(u_1\right) K\left(u_2\right)}{1/\left(1+\exp\left(-\beta_0^0-\beta_1^0 Y_{i1}-\beta_2^0 Y_{i2}\right)\right)} f\left(Y_{i1}, y_{j2}\right) \\
&\quad \times du_2 du_1 dy_{j2} - \int e_1^2\left(Y_{i1}\right) g\left(Y_{i1}\right) K\left(u_1\right) f_1\left(Y_{i1}\right) du_1 \\
&= e_1^2\left(Y_{i1}\right) g\left(Y_{i1}\right) W_i f_1\left(Y_{i1}\right)\left(1+\exp\left(-\beta_0^0-\beta_1^0 Y_{i1}-\beta_2^0 Y_{i2}\right)\right) - e_1^2\left(Y_{i1}\right) g\left(Y_{i1}\right) f_1\left(Y_{i1}\right).
\end{aligned}$$

Then $\Sigma_1 = Var\left[h_1\left(\mathbf{X}\right)\right] = \left\{\sigma_{1,ij}\right\}_{i,j=1}^3$, where

$$\begin{aligned}
\sigma_{1,ij} =& E\left[e_1^4\left(Y_1\right) g_i\left(Y_1\right) g_j\left(Y_1\right) f_1^2\left(Y_1\right)\left(1+\exp\left(-\beta_0^0-\beta_1^0 Y_1-\beta_2^0 Y_2\right)\right)\right] \\
& - E\left[e_1^4\left(Y_1\right) g_i\left(Y_1\right) g_j\left(Y_1\right) f_1^2\left(Y_1\right)\right] \\
=& E\left[e_1^4\left(Y_1\right) g_i\left(Y_1\right) g_j\left(Y_1\right) f_1^2\left(Y_1\right) \exp\left(-\beta_0^0-\beta_1^0 Y_1-\beta_2^0 Y_2\right)\right].
\end{aligned}$$

By the relationship between V- and U-statistics introduced in Section 5.7.3 and Theorem A in Section 5.5.1 Serfling (2009), one has asymptotic normality of $\varphi_N\left(\beta^0\right)$ as

$$\sqrt{N}\varphi_N\left(\beta^0\right) \sim N\left(\mathbf{0}, 4\Sigma_1\right). \qquad \square$$

Lemma 7. *The probability limit of the second derivative of* $M_N\left(\beta\right)$ *is*

$$E\left[\frac{\partial^2}{\partial\beta^2} M_N\left(\beta^0\right)\right] = 2E\left[e_1^2\left(Y_1\right) g\left(Y_1\right) g\left(Y_1\right)^T\right] + o(1).$$

Proof of Lemma 7. Notice that for any $\beta \in \Theta$, one has

$$\begin{aligned}\frac{\partial^2}{\partial\beta^2} M_N(\beta) &= \frac{2}{N}\sum_{i=1}^{N}\left\{e_1^2(Y_{i1})\left[\frac{\partial}{\partial\beta}A_{1\beta}(Y_{i1}) + \frac{\partial}{\partial\beta}B_{1\beta}(Y_{i1})\right]\left[\frac{\partial}{\partial\beta}A_{1\beta}(Y_{i1}) + \frac{\partial}{\partial\beta}B_{1\beta}(Y_{i1})\right]^T\right.\\ &\quad \left. + e_1^2(Y_{i1})\left[A_{1\beta}(Y_{i1}) + B_{1\beta}(Y_{i1}) + C_1(Y_{i1})\right]\left[\frac{\partial^2}{\partial\beta^2}A_{1\beta}(Y_{i1}) + \frac{\partial^2}{\partial\beta^2}B_{1\beta}(Y_{i1})\right]\right\}\\ &= \frac{2}{N}\sum_{i=1}^{N} e_1^2(Y_{i1})\frac{\partial}{\partial\beta}A_{1\beta^0}(Y_{i1})\frac{\partial}{\partial\beta}A_{1\beta^0}(Y_{i1})^T + o_p(1),\end{aligned}$$

due to the fact that $B_{1\beta}(Y_{i1})$, $C_1(Y_{i1})$ and $\frac{\partial}{\partial\beta}B_{1\beta}(Y_{i1})$ are $o_p(1)$. Lemma follows directly by taking expectations on both sides and noting that $A_{1\beta^0}(Y_{i1}) = 0$. □

Second Part, $M_n(\beta)$

Using the notation in (S3.4), the second part of $M_{N,n}(\beta)$ is

$$M_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{e_2(Y_{i2}^r)\left[A_{2\beta}(Y_{i2}^r) + B_{2\beta}(Y_{i2}^r) + C_2(Y_{i2}^r)\right]\right\}^2.$$

The first order derivative of $M_n(\beta)$ is

$$\begin{aligned}\varphi_n(\beta) &= \frac{\partial}{\partial\beta}M_n(\beta)\\ &= \frac{2}{n}\sum_{i=1}^{n} e_2^2(Y_{i2}^r)\left[A_{2\beta}(Y_{i2}^r) + B_{2\beta}(Y_{i2}^r) + C_2(Y_{i2}^r)\right]\left[\frac{\partial}{\partial\beta}A_{2\beta}(Y_{i2}^r) + \frac{\partial}{\partial\beta}B_{2\beta}(Y_{i2}^r)\right].\end{aligned}$$

When $\varphi_n(\beta)$ is evaluated at β^0, $f_2(y_2 \mid \beta^0) = f_2(y_2)$, and $A_{2\beta^0}(y_2) = 0$. Then

$$\varphi_n(\beta^0) = \frac{1}{n}\sum_{i=1}^{n}\left[2e_2^2(Y_{i2}^r)\frac{\partial}{\partial\beta}A_{2\beta^0}(Y_{i2}^r)\left(B_{2\beta^0}(Y_{i2}^r) + C_2(Y_{i2}^r)\right)\right](1 + o_p(1)).$$

Here $\frac{\partial}{\partial\beta}A_{2\beta^0}(y_2) = \mathbf{k}(y_2) = [k_1(y_2), k_2(y_2), k_3(y_2)]^T$ with

$$k_1(y_2) = -\int f(y_1, y_2) / \left\{1 + \exp\left(\beta_0^0 + \beta_1^0 y_1 + \beta_2^0 y_2\right)\right\} dy_1,$$
$$k_2(y_2) = -\int y_1 f(y_1, y_2) / \left\{1 + \exp\left(\beta_0^0 + \beta_1^0 y_1 + \beta_2^0 y_2\right)\right\} dy_1,$$
$$k_3(y_2) = -\int y_2 f(y_1, y_2) / \left\{1 + \exp\left(\beta_0^0 + \beta_1^0 y_1 + \beta_2^0 y_2\right)\right\} dy_1.$$

Define a function $T_2(x, y, z, w) = \int w K_{h_1}(y_1 - x) K_{h_2}(z - y)(1 + \exp(-\beta_0^0 - \beta_1^0 y_1 - \beta_2^0 z)) dy_1 - f_2(z)$. Then we have

$$\begin{aligned}\varphi_n\left(\beta^0\right) =& \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\left[2e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)T_2\left(Y_{j1}, Y_{j2}, Y_{i2}^r, W_j\right)\right] \\ &+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\left[2e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)\left(f_2(Y_{i2}^r) - K_{h_2}\left(Y_{i2}^r - Y_{l2}^r\right)\right)\right] + o_p(1) \\ =& \varphi_{n1}\left(\beta^0\right) + \varphi_{n2}\left(\beta^0\right) + o_p(1).\end{aligned}$$

Lemma 8. *Define* $h_{11}(\mathbf{X}_j) = E\left[2e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)T_2\left(Y_{j1}, Y_{j2}, Y_{i2}^r, W_j\right) \mid \mathbf{X}_j\right]$ *and* $\Sigma_{21} = Var\left[h_{11}(\mathbf{X})\right]$. *Then* $\sqrt{N}\varphi_{n1}(\beta^0) \sim N(\mathbf{0}, \Sigma_{21})$.

Proof of Lemma 8. One notes that

$$\begin{aligned}h_{11}(\mathbf{X}_j) &= E\left[2e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)T_2\left(Y_{j1}, Y_{j2}, Y_{i2}^r, W_j\right) \mid \mathbf{X}_j\right] \\ &= \int 2e_2^2(y_{i2})\mathbf{k}(y_{i2})\int\frac{W_j K_{h_1}(y_1 - Y_{j1}) K_{h_2}(y_{i2} - Y_{j2})}{1/\left(1 + \exp\left(-\beta_0^0 - \beta_1^0 y_1 - \beta_2^0 y_{i2}\right)\right)}dy_1 f_2(y_{i2})\, dy_{i2} \\ &\quad - E\left[2e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)f_2(Y_{i2}^r)\right].\end{aligned}$$

Let $u_1 = \frac{y_1 - Y_{j1}}{h_1}$ and $u_2 = \frac{y_{i2} - Y_{j2}}{h_2}$. With a change of variable and Taylor expansion, one has

$$h_{11}(\mathbf{X}_j) \approx \int 2e_2^2(Y_{j2})\,\mathbf{k}(Y_{j2}) \int \frac{W_j K(u_1) K(u_2)}{1/\left(1+\exp\left(-\beta_0^0 - \beta_1^0 Y_{j1} - \beta_2^0 Y_{j2}\right)\right)} du_1 f_2(Y_{j2})\, du_2$$
$$- E\left[2e_2^2(Y_{i2}^r)\,\mathbf{k}(Y_{i2}^r)\, f_2(Y_{i2}^r)\right]$$

$$= 2e_2^2(Y_{j2})\,\mathbf{k}(Y_{j2})\, W_j \left(1+\exp\left(-\beta_0^0 - \beta_1^0 Y_{j1} - \beta_2^0 Y_{j2}\right)\right) f_2(Y_{j2})$$
$$- 2E\left[e_2^2(Y_{i2}^r)\,\mathbf{k}(Y_{i2}^r)\, f_2(Y_{i2}^r)\right].$$

Then $\Sigma_{21} = Var\left[h_{11}(\mathbf{X})\right] = \{\sigma_{21,ij}\}_{i,j=1}^{3}$, where the ij-th element,

$$\sigma_{21,ij} = 4E\left[e_2^4(Y_2)\, k_i(Y_2)\, k_j(Y_2) \left(1+\exp\left(-\beta_0^0 - \beta_1^0 Y_1 - \beta_2^0 Y_2\right)\right) f_2^2(Y_2)\right]$$
$$- 4E\left[e_2^2(Y_2)\, k_i(Y_2)\, f_2(Y_2)\right] E\left[e_2^2(Y_2)\, k_j(Y_2)\, f_2(Y_2)\right]$$
$$- 4E\left[e_2^2(Y_2)\, k_j(Y_2)\, f_2(Y_2)\right] E\left[e_2^2(Y_2)\, k_i(Y_2)\, f_2(Y_2)\right]$$
$$+ 4E\left[e_2^2(Y_2)\, k_i(Y_2)\, f_2(Y_2)\right] E\left[e_2^2(Y_2)\, k_j(Y_2)\, f_2(Y_2)\right]$$
$$= 4E\left[e_2^4(Y_2)\, k_i(Y_2)\, k_j(Y_2) \left(1+\exp\left(-\beta_0^0 - \beta_1^0 Y_1 - \beta_2^0 Y_2\right)\right) f_2^2(Y_2)\right]$$
$$- 4E\left[e_2^2(Y_2)\, k_i(Y_2)\, f_2(Y_2)\right] E\left[e_2^2(Y_2)\, k_j(Y_2)\, f_2(Y_2)\right].$$

Define $U_N^* = \frac{1}{N}\sum_{j=1}^{N} h_{11}(\mathbf{X}_j)$. By Central Limit Theorem,

$$\sqrt{N} U_N^* \xrightarrow{d} N(\mathbf{0}, \Sigma_{21}).$$

To prove Lemma 8, it is sufficient to show that $\sqrt{N}\left[\varphi_{n1}\left(\beta^0\right)-U_N^*\right]=o_p(1)$. Note that

$$
\begin{aligned}
&Var\left[\sqrt{N}\varphi_{n1}\left(\beta^0\right)\right]=\frac{1}{n^2N}\sum_{i=1}^{n}\sum_{j=1}^{N}Var\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j)\right]\\
&+\frac{1}{n^2N}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{j\neq j'}Cov\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j),2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j'1},Y_{j'2},Y_{i2}^r,W_{j'})\right]\\
&+\frac{1}{n^2N}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{i\neq i'}Cov\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j),2e_2^2(Y_{i'2}^r)\mathbf{k}(Y_{i'2}^r)T_2(Y_{j1},Y_{j2},Y_{i'2}^r,W_j)\right]\\
&\approx\frac{1}{n}Var\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j)\right]+\frac{(n-1)}{n}\Sigma_{21}\approx\Sigma_{21}. \qquad \text{(S3.5)}
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&Cov\left[\sqrt{N}\varphi_{n1}\left(\beta^0\right),\sqrt{N}U_N^*\right]\\
=\;&\frac{1}{nN}Cov\left[\sum_{i=1}^{n}\sum_{j=1}^{N}\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j)\right],\sum_{j'=1}^{N}h_{11}\left(\mathbf{X}_{j'}\right)\right]\\
=\;&\frac{1}{nN}\sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{j=j'}Cov\left[2e_2^2(Y_{i2}^r)\mathbf{k}(Y_{i2}^r)T_2(Y_{j1},Y_{j2},Y_{i2}^r,W_j),h_{11}\left(\mathbf{X}_{j'}\right)\right]\\
=\;&\Sigma_{21}. \qquad \text{(S3.6)}
\end{aligned}
$$

Since $Var\left(\sqrt{N}U_N^*\right)\approx\Sigma_{21}$, together with (S3.5) and (S3.6),

$$
\begin{aligned}
Var\left[\sqrt{N}\varphi_{n1}\left(\beta^0\right)-\sqrt{N}U_N^*\right]\;=\;&Var\left[\sqrt{N}\varphi_{n1}\left(\beta^0\right)\right]+Var\left[\sqrt{N}U_N^*\right]\\
&-2Cov\left[\sqrt{N}\varphi_{n1}\left(\beta^0\right),\sqrt{N}U_N^*\right]=o(1).
\end{aligned}
$$

Therefore, $\sqrt{N}\varphi_{n1}\left(\beta^0\right)-\sqrt{N}U_N^* = o_p(1)$, and $\sqrt{N}\varphi_{n1}\left(\beta^0\right) \sim N\left(\mathbf{0}, \Sigma_{21}\right)$. □

Note that $\varphi_{n2}\left(\beta^0\right)$ is a V-statistic. Let

$$H\left(y_{i2}, y_{l2}\right) = e_2^2\left(y_{i2}\right)\mathbf{k}\left(y_{i2}\right)\left(f_2(y_{i2}) - K_{h_2}\left(y_{i2}-y_{l2}\right)\right) + e_2^2\left(y_{l2}\right)\mathbf{k}\left(y_{l2}\right)\left(f_2(y_{l2}) - K_{h_2}\left(y_{l2}-y_{i2}\right)\right),$$

where y_{i2} and y_{j2} represent independent refreshment samples. Then

$$\varphi_{n2}\left(\beta^0\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n} H\left(Y_{i2}^r, Y_{l2}^r\right). \tag{S3.7}$$

Lemma 9. *Define* $h_{12}\left(Y_{i2}^r\right) = E\left[H\left(Y_{i2}^r, Y_{l2}^r\right) \mid Y_{i2}^r\right]$ *and* $\Sigma_{22} = Var\left[h_{12}\left(Y_{i2}^r\right)\right]$.

Then $E\left[H\left(Y_{i2}^r, Y_{l2}^r\right)\right] = o(1)$ *and* $\sqrt{n}\varphi_{n2}\left(\beta^0\right) \sim N\left(\mathbf{0}, 4\Sigma_{22}\right)$.

Proof of Lemma 9. We have

$$\begin{aligned}
& E\left[H\left(Y_{i2}^r, Y_{l2}^r\right)\right] \\
= \; & E\left[e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)\left[f_2(Y_{i2}^r) - \int K_{h_2}\left(Y_{i2}^r - Y_{l2}^r\right) f_2(y_{l2}) dy_{l2}\right]\right] \\
& + E\left[e_2^2\left(Y_{l2}\right)\mathbf{k}\left(Y_{l2}^r\right)\left[f_2(Y_{l2}^r) - \int K_{h_2}\left(Y_{l2}^r - y_{i2}\right) f_2(y_{i2}) dy_{i2}\right]\right] \\
= \; & E\left[e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right)\left[f_2(Y_{i2}^r) - f_2(Y_{i2}^r)\right]\right] + E\left[e_2^2\left(Y_{l2}^r\right)\mathbf{k}\left(Y_{l2}^r\right)\left[f_2(Y_{l2}^r) - f_2(Y_{l2}^r)\right]\right] + o(1) \\
= \; & o(1).
\end{aligned}$$

Then $h_{12}\left(Y_{i2}^r\right) = \left\{E\left[e_2^2\left(Y_{l2}^r\right)\mathbf{k}\left(Y_{l2}^r\right) f_2(Y_{l2}^r)\right] - e_2^2\left(Y_{i2}^r\right)\mathbf{k}\left(Y_{i2}^r\right) f_2\left(Y_{i2}^r\right)\right\}(1+o(1))$, and define $\Sigma_{22} = Var\left[h_{12}\left(Y_2\right)\right] = \left\{\sigma_{22,ij}\right\}_{i,j=1}^{3}$ with its ij-th element being

$$\begin{aligned}\sigma_{22,ij} &= E\left[e_2^4\left(Y_2\right)k_i\left(Y_2\right)k_j\left(Y_2\right)f_2^2\left(Y_2\right)\right] \\ &\quad -E\left[e_2^2\left(Y_2\right)k_i\left(Y_2\right)f_2(Y_2)\right]E\left[e_2^2\left(Y_2\right)k_j\left(Y_2\right)f_2(Y_2)\right].\end{aligned}$$

By the property of V-statistics, one has $\sqrt{n}\varphi_{n2}\left(\beta^0\right) \sim N\left(\mathbf{0}, 4\Sigma_{22}\right)$. □

Lemma 10. *The probability limit of the second derivative of $M_n\left(\beta\right)$ is*

$$E\left[\frac{\partial^2}{\partial\beta^2}M_n\left(\beta^0\right)\right] = 2E\left[e_2^2\left(Y_2\right)\mathbf{k}\left(Y_2\right)\mathbf{k}\left(Y_2\right)^T\right] + o(1).$$

Proof of Lemma 10. One notes that

$$\begin{aligned}&\frac{\partial^2}{\partial\beta^2}M_n\left(\beta\right) \\ =\; &\frac{2}{n}\sum_{i=1}^{n}\left\{e_2^2\left(Y_{i2}^r\right)\left[\frac{\partial}{\partial\beta}A_{2\beta}(Y_{i2}^r) + \frac{\partial}{\partial\beta}B_{2\beta}(Y_{i2}^r)\right]\left[\frac{\partial}{\partial\beta}A_{2\beta}(Y_{i2}^r) + \frac{\partial}{\partial\beta}B_{2\beta}(Y_{i2}^r)\right]^T\right. \\ &\left. + e_2^2\left(Y_{i2}^r\right)\left[A_{2\beta}(Y_{i2}^r) + B_{2\beta}(Y_{i2}^r) + C_2(Y_{i2}^r)\right]\left[\frac{\partial^2}{\partial\beta^2}A_{2\beta}(Y_{i2}^r) + \frac{\partial^2}{\partial\beta^2}B_{2\beta}(Y_{i2}^r)\right]\right\}.\end{aligned}$$

Therefore, $\frac{\partial^2}{\partial\beta^2}M_n\left(\beta^0\right) = \frac{2}{n}\sum_{i=1}^{N}e_2^2\left(Y_{i2}^r\right)\frac{\partial}{\partial\beta}A_{2\beta^0}(Y_{i2}^r)\frac{\partial}{\partial\beta}A_{2\beta^0}(Y_{i2}^r)^T + o_p(1)$. It is due to the fact that $B_{2\beta}(Y_{i2}^r)$, $C_2(Y_{i2}^r)$ and $\frac{\partial}{\partial\beta}B_{2\beta}(Y_{i2}^r)$ are $o_p\left(1\right)$. Then the result follows directly by taking expectations. □

Proof of Theorem 3. Now we have three multivariate normal distributed vectors, namely $\varphi_N\left(\beta^0\right)$, $\varphi_{n1}\left(\beta^0\right)$ and $\varphi_{n2}\left(\beta^0\right)$. The sum of these vectors is again a multivariate normal distributed vector. By the relationship between V-statistics and U-statistics and the proof of asymptotic properties of U-statistics, we can rewrite these three random vectors as

$$\begin{aligned}\varphi_N\left(\beta^0\right) &\approx \widetilde{\varphi}_N\left(\beta^0\right)=\frac{2}{N}\sum_{i=1}^{N}h_1\left(\mathbf{X}_i\right)\sim N\left(\mathbf{0},\frac{4}{N}\Sigma_1\right),\\ \varphi_{n1}\left(\beta^0\right) &\approx \widetilde{\varphi}_{n1}\left(\beta^0\right)=\frac{1}{N}\sum_{j=1}^{N}h_{11}\left(\mathbf{X}_j\right)\sim N\left(\mathbf{0},\frac{1}{N}\Sigma_{21}\right),\\ \varphi_{n2}\left(\beta^0\right) &\approx \widetilde{\varphi}_{n2}\left(\beta^0\right)=\frac{2}{n}\sum_{l=1}^{n}h_{12}\left(Y_{l2}^r\right)\sim N\left(\mathbf{0},\frac{4}{n}\Sigma_{22}\right),\end{aligned} \tag{S3.8}$$

where $\mathbf{X}_i$ and $\mathbf{X}_j$ represent the sample from the panel and Y_{l2}^r the sample from refreshment, therefore they are independent. As a result

$$Cov\left[h_1\left(\mathbf{X}_i\right),h_{12}\left(Y_{l2}^r\right)\right]=0 \text{ and } Cov\left[h_{11}\left(\mathbf{X}_j\right),h_{12}\left(Y_{l2}^r\right)\right]=0.$$

The covariance contribution is between $\varphi_N\left(\beta^0\right)$ and $\varphi_{n1}\left(\beta^0\right)$. For $i\neq j$, we have $Cov\left[h_1\left(\mathbf{X}_i\right),h_{11}\left(\mathbf{X}_j\right)\right]=0$. And for $i=j$, we have

$$\begin{aligned}\Sigma_{cov} &= Cov\left[h_1\left(\mathbf{X}\right),h_{11}\left(\mathbf{X}\right)\right]=E\left[h_1\left(\mathbf{X}\right)h_{11}\left(\mathbf{X}\right)^T\right]\\ &= \left\{E\left[h_{ij}^{cov}\right]\right\}_{i,j=1}^{3},\end{aligned} \tag{S3.9}$$

where h_{ij}^{cov} is the ij^{th} element of matrix $h_1(x)\,h_{11}^T(x)$ and

$$\begin{aligned}
h_{ij}^{cov} = &\left[e_1^2(y_1)\,g_i(y_1)\,wf_1(y_1)\left(1+\exp\left(-\beta_0^0-\beta_1^0y_1-\beta_2^0y_2\right)\right)-e_1^2(y_1)\,g_i(y_1)\,f_1(y_1)\right]\\
&\times\left[2e_2^2(y_2)\,k_j(y_2)\,w\left(1+\exp\left(-\beta_0^0-\beta_1^0y_1-\beta_2^0y_2\right)\right)f_2(y_2)\right.\\
&\left.-2E\left[e_2^2(Y_2)\,k_j(Y_2)\,f_2(Y_2)\right]\right]\\
=&2e_1^2(y_1)\,e_2^2(y_2)\,g_i(y_1)\,k_j(y_2)\,w^2\left(1+\exp\left(-\beta_0^0-\beta_1^0y_1-\beta_2^0y_2\right)\right)^2 f_1(y_1)\,f_2(y_2)\\
&-2e_1^2(y_1)\,g_i(y_1)\,wf_1(y_1)\left(1+\exp\left(-\beta_0^0-\beta_1^0y_1-\beta_2^0y_2\right)\right)E\left[e_2^2(Y_2)\,k_j(Y_2)\,f_2(Y_2)\right]\\
&-2e_1^2(y_1)\,e_2^2(y_2)\,g_i(y_1)\,k_j(y_2)\,w\left(1+\exp\left(-\beta_0^0-\beta_1^0y_1-\beta_2^0y_2\right)\right)f_1(y_1)\,f_2(y_2)\\
&+2e_1^2(y_1)\,g_i(y_1)\,f_1(y_1)\,E\left[e_2^2(Y_2)\,k_j(Y_2)\,f_2(Y_2)\right].
\end{aligned}$$

Then

$$\begin{aligned}
E\left[h_{ij}^{cov}\right]=&2E\left[e_1^2(Y_1)\,e_2^2(Y_2)\,g_i(Y_1)\,k_j(Y_2)\left(1+\exp\left(-\beta_0^0-\beta_1^0Y_1-\beta_2^0Y_2\right)\right)f_1(Y_1)\,f_2(Y_2)\right]\\
&-2E\left[e_1^2(Y_1)\,e_2^2(Y_2)\,g_i(Y_1)\,k_j(Y_2)\,f_1(Y_1)\,f_2(Y_2)\right]\\
=&2E\left[e_1^2(Y_1)\,e_2^2(Y_2)\,g_i(Y_1)\,k_j(Y_2)\,f_1(Y_1)\,f_2(Y_2)\exp\left(-\beta_0^0-\beta_1^0Y_1-\beta_2^0Y_2\right)\right].
\end{aligned}$$

Let $N=rn$, r is the ratio between N and n. Then we have

$$\sqrt{N}\left[\varphi_N\left(\beta^0\right)+\varphi_{n1}\left(\beta^0\right)+\varphi_{n2}\left(\beta^0\right)\right]\sim N\left(\mathbf{0},4\Sigma_1+\Sigma_{21}+4r\Sigma_{22}+4\Sigma_{cov}\right).$$

Define $\Sigma=4\Sigma_1+\Sigma_{21}+4r\Sigma_{22}+4\Sigma_{cov}$ and $V=E\left[\frac{\partial^2}{\partial\beta^2}M_N\left(\beta^0\right)\right]+E\left[\frac{\partial^2}{\partial\beta^2}M_n\left(\beta^0\right)\right]$. By Theorem 5.21 Van der Vaart (2000) we have the asymptotic property for $\widehat{\beta}$

as follow

$$\sqrt{N}\left(\widehat{\beta}-\widehat{\beta}_0\right) \sim N\left(\mathbf{0}, \left(V^{-1}\right)\Sigma\left(V^{-1}\right)^T\right). \qquad \square$$

Bibliography

Devroye, L. P. and T. J. Wagner (1980). The strong uniform consistency of kernel density estimates. In Multivariate Analysis V: Proceedings of the fifth International Symposium on Multivariate Analysis, Volume 5, pp. 59–77.

Hirano, K., G. W. Imbens, G. Ridder, and D. B. Rubin (2001). Combining panel data sets with attrition and refreshment samples. Econometrica 69(6), 1645–1659.

Serfling, R. J. (2009). Approximation theorems of mathematical statistics, Volume 162. John Wiley & Sons.

Silverman, B. W. et al. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. The Annals of Statistics 6(1), 177–184.

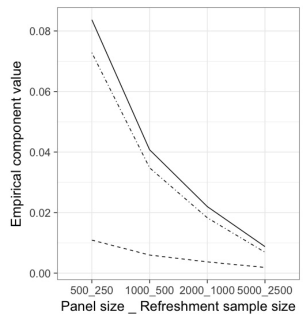
Van der Vaart, A. W. (2000). Asymptotic statistics. Cambridge university press.

	Obs	Y_1	Y_2	X	W
Complete set	1	Y_{11}	Y_{12}	X_1	W_1=1
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	n_c	Y_{n_c1}	Y_{n_c2}	X_{n_c}	W_c=1
Incomplete set	n_c+1	$Y_{(n_c+1)1}$		X_{n_c+1}	W_{n_c+1}=0
	$\vdots$	$\vdots$		$\vdots$	$\vdots$
	N	Y_{N1}		X_N	W_N=0
Refreshment sample	1		Y^r_{12}	X^r_1	
	$\vdots$		$\vdots$	$\vdots$	
	n		Y^r_{n2}	X^r_n	

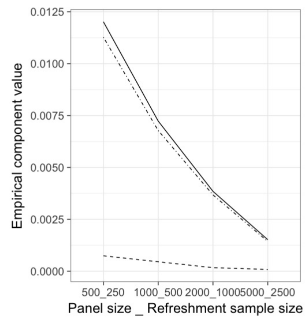
Table 1: Observed full data set with one categorical explanatory variable.

		Asymptotic Formula		Simulation	
σ_1^2	σ_2^2	$SE_{\widehat{\beta}_1}$	$SE_{\widehat{\beta}_2}$	$SE_{\widehat{\beta}_1}$	$SE_{\widehat{\beta}_2}$
1	1	0.105	0.161	0.081	0.131
5	5	0.043	0.066	0.036	0.055
10	10	0.035	0.048	0.028	0.041

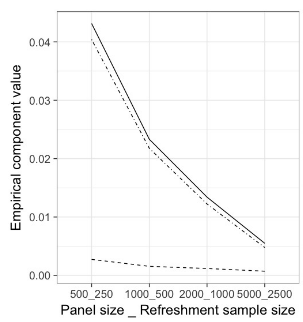
Table 2: The effect of marginal variances σ_1^2 and σ_2^2 on standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ computed from both the asymptotic formula and simulation. The panel size is 5000 and the refreshment sample size is 2500. True values of attrition parameters β_1 and β_2 are 0.3 and 0.4 respectively.



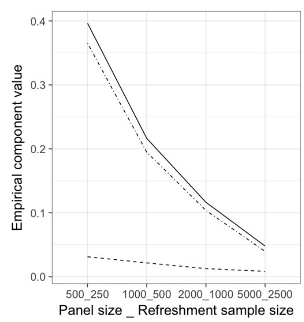
(a) $\widehat{\beta}_0$



(b) $\widehat{\beta}_1$



(c) $\widehat{\beta}_2$



(d) $\widehat{\beta}_3$

Figure 1: Large sample performance of semi-parametric estimators in the one-covariate case. The dashed, dot-dash, and solid lines represent the squared bias, variance, and MSE respectively.

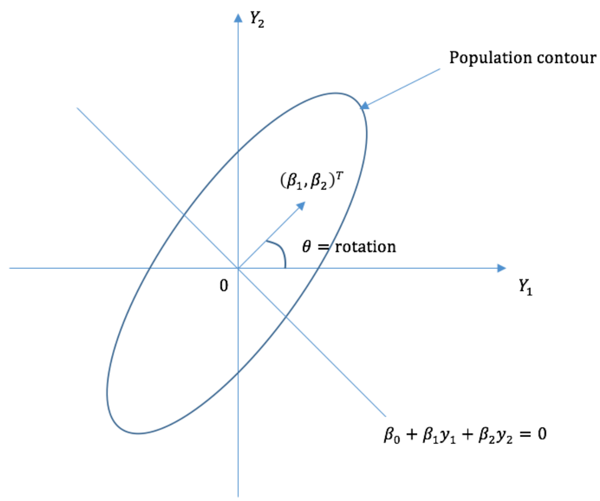


Figure 2: The definitions of the reference line, perpendicular vector, and rotation angle. Here the contour of population joint density of (Y_1, Y_2) is also plotted.

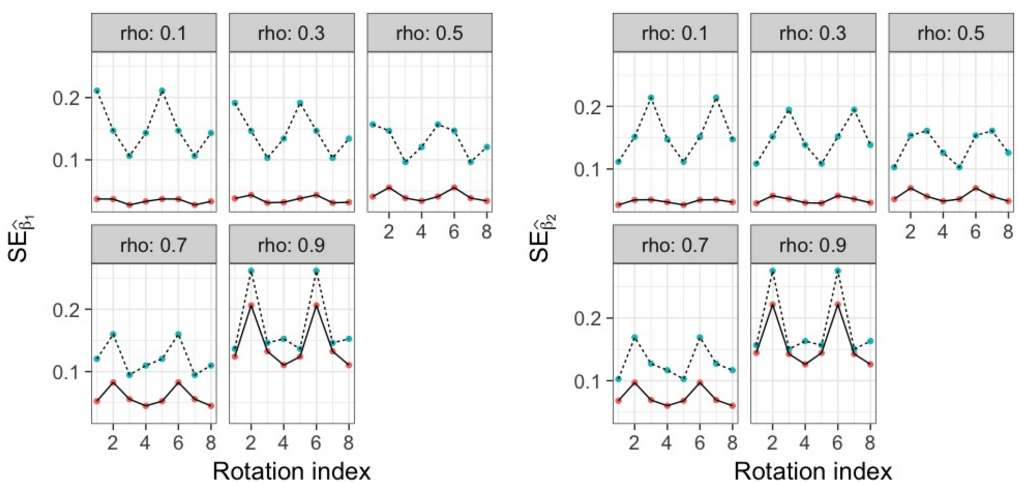


Figure 3: The effect of the perpendicular vector (PV) rotation on standard errors of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ computed from the asymptotic formula. The solid and dashed lines are for PV lengths of 0.5 and 1 respectively. Five levels of correlation coefficient are considered with values $0.1, 0.3, 0.5, 0.7$ and 0.9. The panel size is 5000 and the refreshment sample size is 2500.