Supplementary Material for "Gaussian variational approximation with composite likelihood for crossed random effect models"

Libai Xu, Nancy Reid and Dehan Kong

Soochow University and University of Toronto

The supplementary material contains the notation and assumptions in §S1, proofs of the main theorems for the Poisson regression model in §S2 and §S3, proofs of the main theorems for the Gamma regression model in §S4 and §S5, additional detailed simulation results in §S6, and additional details on real data analysis in §S7.

S1 Notation and Assumptions

First, recall the Poisson regression model with crossed random effects,

$$Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Poisson}(\mu_{ij}), \mu_{ij} = \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j), \quad \text{(S1.1)}$$

$$U_i \sim N(0, (\sigma_u^2)^0), V_j \sim N(0, (\sigma_v^2)^0), U_i \text{ and } V_j \text{ are independent.}$$

and the Gamma regression model with crossed random effects,

$$
\begin{aligned}
&Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Gamma}(\alpha, \mu_{ij}), \\
&\mu_{ij} = \exp(\mathbf{X}_{ij}\boldsymbol{\beta}^0 + U_i + V_j), \qquad\qquad (\text{S1.2}) \\
&U_i \sim N(0, (\sigma_u^2)^0), V_j \sim N(0, (\sigma_v^2)^0), U_i \text{ and } V_j \text{ are independent,}
\end{aligned}
$$

where the shape parameter α is assumed known.

S1.1 Assumptions

Our proofs rely on the following assumptions:

- (A1) The sets of variables $\mathcal{S}_i = \{(\mathbf{X}_{ij}, Y_{ij}, U_i, V_j) : j = 1, ..., n\}$, for $i = 1, ..., m$, are conditional independent on $\{V_1, ..., V_n\}$ and identically distributed, and for any given V_j, the elements $(\mathbf{X}_{ij}, Y_{ij}, U_i), i = 1, ..., m$, are identically distributed as $(\mathbf{X}, Y, U)$.

- (A2) The sets of variables $\mathcal{S}'_j = \{(\mathbf{X}_{ij}, Y_{ij}, U_i, V_j) : i = 1, ..., m\}$, for $j = 1, ..., n$, are conditional independent on $\{U_1, ..., U_m\}$ and identically distributed, and for any given U_i, the elements $(\mathbf{X}_{ij}, Y_{ij}, V_j), j = 1, ..., n$, are identically distributed as $(\mathbf{X}, Y, V)$.

- (A3) The quadruples $(\mathbf{X}_{ij},Y_{ij},U_i,V_j)$ and $(\mathbf{X}_{kl}, Y_{kl}, U_k, V_l)$ for $k \neq i, l \neq j$ are multually independent and identically distributed as (X, Y, U, V).

- (A4) The random variables $\mathbf{X}$, U, and V are mutually independent.

- (A5) The true values $\boldsymbol{\beta}^0, (\sigma_u^2)^0$, and $(\sigma_v^2)^0$ of $\boldsymbol{\beta}, \sigma_u^2$, and σ_v^2, respectively, lie in $\mathbb{R}^p, (0, \infty)$, and $(0, \infty)$; when choosing $(\underline{\widehat{\boldsymbol{\beta}}}, \underline{\widehat{\sigma}}_u^2, \underline{\widehat{\sigma}}_v^2, \underline{\widehat{\mu}}_u, \underline{\widehat{\lambda}}_u, \underline{\widehat{\mu}}_v, \underline{\widehat{\lambda}}_v)$ we search in the rectangular region $[-C_1, C_1]^p \times [C_1^{-1}, C_1] \times [C_1^{-1}, C_1] \times [-C_2, C_2]^m \times [-C_1, C_1]^m \times [-C_2, C_2]^n \times [-C_1, C_1]^n$, where C_1 is a constant satisfying $C_1 > \max\{|\beta_0^0|, ..., |\beta_p^0|, (\sigma_u^2)^0, 1/(\sigma_u^2)^0, (\sigma_v^2)^0, 1/(\sigma_v^2)^0\}$ and C_2 is a sufficiently large but finite positive constant.

- (A6) The moment generating function of $\mathbf{X} = (X_1, ..., X_p)^{\mathrm{T}}$, $\phi(t) = E\{\exp(\mathbf{X}^{\mathrm{T}} t)\}$ is well defined in $\mathbb{R}^p$.

- (A7) The functions

$$\begin{aligned} \phi_1(t) &= E_{\mathbf{X}}\{\mathbf{X} \exp(\mathbf{X}^{\mathrm{T}} t)\}, \\ \phi_2(t) &= E_{\mathbf{X}}\{\mathbf{X}\mathbf{X}^{\mathrm{T}} \exp(\mathbf{X}^{\mathrm{T}} t)\}, \end{aligned}$$

  are well defined for all vector t. For simplicity, we denote $\phi(\cdot) = \phi_0(\cdot)$;

- (A8) In some neighborhood of $\boldsymbol{\beta}_{-1}^0 = (\beta_2^0, ..., \beta_p^0)^{\mathrm{T}}$, $(d^2/d\beta_i^2) \log\{\phi(\beta_i)\}$ does not vanish, $i = 1, ..., p$;

- (A9) The mapping that takes $\boldsymbol{\beta}$ to $\phi_1(\boldsymbol{\beta})/\phi(\boldsymbol{\beta})$ is invertible.

The assumptions (A6), (A8) and (A9) were used explicitly by Hall et al. (2011), whilst (A1)-(A5) are the extensions of (A1)-(A3) and (A11) of Hall et al. (2011). Assumption (A7) was used to make the results representation concise by Ormerod and Wand (2012).

S1.2 Notation

Denote $\boldsymbol{\Psi}^{rc} = (\beta_1^r, \beta_1^c, \boldsymbol{\beta}_{-1}^{\mathrm{T}}, \sigma_u^2, \sigma_v^2)^{\mathrm{T}}, \boldsymbol{\xi} = (\mu_{u_1}, ..., \mu_{u_m}, \mu_{v_1}, ..., \mu_{v_n}, \lambda_{u_1}, ..., \lambda_{u_m}, \lambda_{v_1}, ..., \lambda_{v_n})^{\mathrm{T}}$; $\widehat{\underline{\beta}_1^r}$ the variational composite likelihood estimate (VCLE) of $\beta_1^{r0} = \beta_1^0 + \sigma_v^2/2$; $\widehat{\underline{\beta}_1^c}$ the VCLE of $\beta_1^{c0} = \beta_1^0 + \sigma_u^2/2$; $\widehat{\underline{\beta}}_1$ the VCLE of intercept β_0^0; $\widehat{\underline{\beta}}_i$ the VCLE of the covariate coefficient $\beta_i^0, i = 2, ..., p$; $\widehat{\underline{\sigma}}_u^2$ the VCLE of variance of random effect $U, (\sigma_u^2)^0$; $\widehat{\underline{\sigma}}_v^2$ the VCLE of variance of random effect $V, (\sigma_v^2)^0$; $\widehat{\underline{\mu}}_{u_i}, \widehat{\underline{\lambda}}_{u_i}$ the VCLEs of parameters of the i-th variational distribution $N(\mu_{u_i}, \lambda_{u_i}), i = 1, ..., m$; $\widehat{\underline{\mu}}_{v_j}, \widehat{\underline{\lambda}}_{v_j}$ the VCLEs of parameters of the j-th variational distribution $N(\mu_{v_j}, \lambda_{v_j}), j = 1, ..., n$.

S2 Proof of Theorem 1 for Poisson regression model

The proof for the consistency and rates of convergence for the estimators $\widehat{\underline{\beta}}_1$, $\widehat{\underline{\boldsymbol{\beta}}}_{-1}$, $\widehat{\underline{\sigma}}_u^2$, and $\widehat{\underline{\sigma}}_v^2$ involves two main steps:

Step1. Establish the consistency of estimators $\beta_1^0, \boldsymbol{\beta}_{-1}^0, (\sigma_u^2)^0$, and $(\sigma_v^2)^0$ in §S2.1;

Step2. Extend the consistency results to a rate of convergence by controlling the remainder terms in §S2.2.

S2.1 Consistency

The variational lower bound of the CL for the Poisson regression model with crossed random effects is

$$
\begin{aligned}
& \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi}) \\
= & \sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} + \mu_{u_i}) - \exp(\mathbf{X}_{ij}\boldsymbol{\beta}^{r} + \mu_{u_i} + \lambda_{u_i}/2)\} \\
& -\frac{m}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}) + \frac{1}{2}\log\lambda_{u_i} \\
& +\sum_{j=1}^{n}\sum_{i=1}^{m}\{Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} + \mu_{v_j}) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} + \mu_{v_j} + \lambda_{v_j}/2)\} \\
& -\frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{j=1}^{n}(\mu_{v_j}^2 + \lambda_{v_j}) + \frac{1}{2}\log\lambda_{v_j}.
\end{aligned}
$$

To derive the consistency of the estimates, we take derivatives of $\underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})$ with respect to the parameters.

$$
\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \beta_1^r} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} + \mu_{u_i} + \lambda_{u_i}/2}\right), \tag{S2.3}
$$

$$
\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \beta_1^c} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} + \mu_{v_j} + \lambda_{v_j}/2}\right), \tag{S2.4}
$$

$$
\begin{aligned}
\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_{-1}} = & \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} + \mu_{u_i} + \lambda_{u_i}/2}\right)\mathbf{X}_{ij,-1} \qquad \text{(S2.5)} \\
& + \sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} + \mu_{v_j} + \lambda_{v_j}/2}\right)\mathbf{X}_{ij,-1},
\end{aligned}
$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \sigma_u^2} = -\frac{m}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}), \tag{S2.6}$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \sigma_v^2} = -\frac{n}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum_{j=1}^{n}(\mu_{v_j}^2 + \lambda_{v_j}), \tag{S2.7}$$

for each μ_{u_i} and λ_{u_i},

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \mu_{u_i}} = \sum_{j=1}^{n}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{u_i} + \lambda_{u_i}/2}\right) - \frac{\mu_{u_i}}{\sigma_u^2}, \tag{S2.8}$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \lambda_{u_i}} = -\frac{1}{2}\sum_{j=1}^{n} e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{u_i} + \lambda_{u_i}/2} - \frac{1}{2\sigma_u^2} + \frac{1}{2\lambda_{u_i}}, \tag{S2.9}$$

for each μ_{v_j} and λ_{v_j},

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \mu_{v_j}} = \sum_{i=1}^{m}\left(Y_{ij} - e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{v_j} + \lambda_{v_j}/2}\right) - \frac{\mu_{v_j}}{\sigma_v^2}, \tag{S2.10}$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \lambda_{v_j}} = -\frac{1}{2}\sum_{i=1}^{m} e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{v_j} + \lambda_{v_j}/2} - \frac{1}{2\sigma_v^2} + \frac{1}{2\lambda_{v_j}}. \tag{S2.11}$$

Setting equations (S2.6) and (S2.7) to zeros and solving we obtain

$$\underline{\widehat{\sigma}}_u^2 = \frac{1}{m}\sum_{i=1}^{m}(\underline{\widehat{\mu}}_{u_i}^2 + \underline{\widehat{\lambda}}_{u_i}), \tag{S2.12}$$

$$\underline{\widehat{\sigma}}_v^2 = \frac{1}{n}\sum_{j=1}^{n}(\underline{\widehat{\mu}}_{v_j}^2 + \underline{\widehat{\lambda}}_{v_j}). \tag{S2.13}$$

Setting (S2.3), (S2.4), (S2.8), and (S2.10) to zeros we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\lambda_{u_i}/2}\right) = 0, \tag{S2.14}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c}+\mu_{v_j}+\lambda_{v_j}/2}\right) = 0, \tag{S2.15}$$

$$\sum_{j=1}^{n}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\lambda_{u_i}/2}\right)-\frac{\mu_{u_i}}{\sigma_u^2} = 0, i=1,...,m, \tag{S2.16}$$

$$\sum_{i=1}^{m}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\lambda_{u_i}/2}\right)-\frac{\mu_{v_j}}{\sigma_v^2} = 0, j=1,...,n. \tag{S2.17}$$

By (S2.16), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\lambda_{u_i}/2}\right)-\sum_{i=1}^{m}\frac{\mu_{u_i}}{\sigma_u^2}=0. \tag{S2.18}$$

Combining (S2.14) and (S2.18) we have

$$\sum_{i=1}^{m}\underline{\widehat{\mu}}_{u_i}=0. \tag{S2.19}$$

By (S2.17), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\left(Y_{ij}-e^{\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\lambda_{u_i}/2}\right)-\sum_{i=1}^{n}\frac{\mu_{v_j}}{\sigma_v^2}=0. \tag{S2.20}$$

Combining (S2.15) and (S2.20) we have

$$\sum_{j=1}^{n}\underline{\widehat{\mu}}_{v_j}=0. \tag{S2.21}$$

Setting equations (S2.9) and (S2.11) to zeros we have $\underline{\widehat{\lambda}}_{u_i}$ and $\underline{\widehat{\lambda}}_{v_j}$ satisfy

$$\underline{\widehat{\lambda}}_{u_i}=\left\{\sigma_u^{-2}+\sum_{j=1}^{n}\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\underline{\widehat{\mu}}_{u_i}+\underline{\widehat{\lambda}}_{u_i}/2)\right\}^{-1}, \tag{S2.22}$$

and

$$\underline{\widehat{\lambda}}_{v_j} = \left\{ \sigma_v^{-2} + \sum_{i=1}^{m} \exp(\mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^c + \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2) \right\}^{-1},$$

from which it is obvious that $0 < \underline{\widehat{\lambda}}_{u_i} < \sigma_u^2$ and $0 < \underline{\widehat{\lambda}}_{v_j} < \sigma_v^2$.

Since $\mathbf{X}$ has a finite moment generating function ϕ on $\mathbb{R}^p$, $\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}$, $\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1})$ is continuous in $[-C_1, C_1]^{p-1}$ with probability one, and $\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) \leq \exp(C_1 \|\mathbf{X}_{ij}\|)$, by the uniform law of large numbers in Jennrich (1969), we have

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right| \to 0,$$

in probability, i.e., for all $\epsilon > 0$, we have

$$\lim_{n \to \infty} P \left\{ \sup_{\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right| > \epsilon \right\} = 0.$$

For the compact interval $[-C_1, C_1]^{p-1}$, we can find a finite sub-cover such that $[-C_1, C_1]^{p-1}$ is covered by $\cup_{k=1}^{K} B(\boldsymbol{\beta}_{-1,k}, \delta)$, and

$$\begin{aligned} & P \left\{ \sup_{\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right| > \epsilon \right\} \qquad \text{(S2.23)} \\ = \; & P \left\{ \sup_{\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q > \epsilon^q \right\} \\ = \; & P \left\{ \max_{k \in [1, \ldots, K]} \sup_{\boldsymbol{\beta}_{-1} \in B(\boldsymbol{\beta}_{-1,k}, \delta)} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q > \epsilon^q \right\}. \end{aligned}$$

In each interval $B(\boldsymbol{\beta}_{-1,k}, \delta)$, we can find a sub-sequence $\mathcal{S}_n^k = \{\boldsymbol{\beta}_{-1,k,l}, l =$

$1, ..., C_0 n^{C'}\}$, C_0 and C' are positive constants, such that

$$\begin{aligned} -\epsilon^q &< \sup_{\boldsymbol{\beta}_{-1} \in B(\boldsymbol{\beta}_{-1,k},\delta)} \left| \frac{1}{n} \sum_{i=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q \\ &\quad - \sup_{\boldsymbol{\beta}_{-1} \in \mathcal{S}_n^k} \left| \frac{1}{n} \sum_{i=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q < \epsilon^q \end{aligned}$$

Then by (S2.23), we have

$$\begin{aligned} & P\left\{ \sup_{\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right| > \epsilon \right\} \\ \leq\ & P\left\{ \max_{k \in \{1,...,K\}} \sup_{\boldsymbol{\beta}_{-1} \in \mathcal{S}_n^k} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q > 2\epsilon^q \right\} \\ \leq\ & P\left[\cup_{k=1}^{K} \cup_{l=1}^{C_0 n^{C'}} \left\{ \omega : \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q > 2\epsilon^q \right\} \right] \\ \leq\ & \sum_{k=1}^{K} \sum_{l=1}^{C_0 n^{C'}} P\left\{ \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q > 2\epsilon^q \right\} \\ \leq\ & 0.5 K C_0 n^{C'} \epsilon^{-q} E \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q . \end{aligned}$$

Since for any $\boldsymbol{\beta}_{-1}$,

$$\frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) = O_p(n^{-1/2}),$$

taking $\epsilon = n^{-\rho}, \rho \in (0, 1/2)$, and by

$$E \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right|^q = O(n^{-q/2}),$$

we have

$$\begin{aligned} & P\left\{ \sup_{\boldsymbol{\beta}_{-1} \in [-C_1, C_1]^{p-1}} \left| \frac{1}{n} \sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1}) \right| > \epsilon \right\} \\ \leq\ & 0.5 K C_0 n^{1+\rho C'} O(n^{-q/2}) = O(n^{-C_2}), \end{aligned}$$

where q satisfies $C_2 = q/2 - (1+q\rho)C' > 0$.

Therefore, under assumption $m = O(n^C)$, we have following properties

$$
\begin{aligned}
& P\left\{\max_{1\le i\le m} \sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\le C_1} \left|\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1})\right| > n^{-\rho}\right\} \\
\le\ & \sum_{i=1}^{c_1 n^C} P\left\{\sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\le C_1} \left|\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1})\right| > n^{-\rho}\right\} \\
=\ & O(n^{-C_2}),
\end{aligned} \tag{S2.24}
$$

where $C_2 = q/2 - C - (1+q\rho)C' > 0$. Since we always can choose a large q such that $C_2 > 0$, without loss of generality, we will assume $C_2 > 0$ changes with lines. By the similar proof process of (S2.24), we derive that

$$
\begin{aligned}
& P\left\{\max_{1\le j\le n} \sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\le C_1} \left|\frac{1}{m}\sum_{i=1}^{m} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) - \phi(\boldsymbol{\beta}_{-1})\right| > m^{-\rho}\right\} \\
=\ & O(m^{-C_2}),
\end{aligned} \tag{S2.25}
$$

The following two properties follow the similar proof process of (S2.24) and (S2.25), respectively,

$$
\begin{aligned}
& P\left\{\max_{1\le i\le m} \sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\le C_1} \left\|\frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}^{\mathrm{T}} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) - \phi_1(\boldsymbol{\beta}_{-1})\right\| > n^{-\rho}\right\} \\
=\ & O(n^{-C_2}).
\end{aligned} \tag{S2.26}
$$

and

$$
\begin{aligned}
& P\left\{\max_{1\le j\le n} \sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\le C_1} \left\|\frac{1}{m}\sum_{i=1}^{m} \mathbf{X}_{ij,-1}^{\mathrm{T}} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) - \phi_1(\boldsymbol{\beta}_{-1})\right\| > m^{-\rho}\right\} \\
=\ & O(m^{-C_2}).
\end{aligned} \tag{S2.27}
$$

Since $\mathbf{X}_{ij,-1}, U_i$, and V_j are mutually independent, $E\{\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} + V_j)\} = \phi(\boldsymbol{\beta}_{-1})\exp\{(\sigma_v^2)^0/2\}$ and $E\{\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}+U_i)\} = \phi(\boldsymbol{\beta}_{-1})\exp\{(\sigma_u^2)^0/2\}$. Then we can derive analogously that

$$P\left\{\max_{1\leq i\leq n}\sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\leq C_1}\left|\frac{1}{n}\sum_{j=1}^{n}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}+V_j)\right.\right.$$
$$\left.\left.-\phi(\boldsymbol{\beta}_{-1})\exp(\sigma_v^2/2)\right| > n^{-\rho}\right\} = O(n^{-C_2}). \tag{S2.28}$$

and

$$P\left\{\max_{1\leq j\leq m}\sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\leq C_1}\left|\frac{1}{m}\sum_{j=1}^{m}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}+U_i)\right.\right.$$
$$\left.\left.-\phi(\boldsymbol{\beta}_{-1})\exp(\sigma_u^2/2)\right| > m^{-\rho}\right\} = O(m^{-C_2}). \tag{S2.29}$$

For the independent normal random variables $U_1, ..., U_m$, they have the following property,

$$P\left\{\max_{1\leq i\leq m} U_i \geq \sigma_u(2D\log m)^{1/2}\right\} \;\leq\; m^{1-D},$$

by assumption $m = O(n^C)$, there exists $C' > 0$ such that

$$P\left[\max_{1\leq i\leq m}\exp(U_i) \geq \exp\{C'(\log n)^{1/2}\}\right] \;\to\; 0. \tag{S2.30}$$

Similarly, by the assumption $n = O(m^{C^*})$, there exists C'' such that

$$P\left[\max_{1\leq j\leq n}\exp(V_j) \geq \exp\{C''(\log m)^{1/2}\}\right] \;\to\; 0. \tag{S2.31}$$

Since for each $1 \leq i \leq m$, conditional on $\mathbf{X}_{ij}, U_i$, and V_j, Y_{ij} follows independent Poisson with mean $\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}+U_i+V_j), 1 \leq j \leq n$. It therefore can be proved using Markov's inequality that there exists a positive constant C_2' and for $\rho \in (0, 1/2)$,

$$\max_{1\le i\le m} P\left[\left|\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}\right| > n^{-\rho}\right] = O(n^{-C_2'}).$$

Hence, by the assumption $m = O(n^C)$,

$$\begin{aligned}
& P\left[\max_{1\le i\le m}\left|\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}\right| > n^{-\rho}\right] \\
= \;& P\left(\cup_{i=1}^{m}\left[\omega : \left|\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}\right| > n^{-\rho}\right]\right) \\
\le \;& \sum_{i=1}^{C_1 n^C} P\left(\left[\omega : \left|\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}\right| > n^{-\rho}\right]\right) \\
\le \;& O(n^{C-C_2'}) = O(n^{-C_2}), \qquad \text{(S2.32)}
\end{aligned}$$

where C_1 is a positive constant. Combining (S2.28), (S2.30), and (S2.32), we can choose C_2' such that $C_2 = C_2' - C > 0$ and for any $\rho \in (0, 1/2)$, we have

$$\begin{aligned}
& P\left\{\max_{1\le i\le m}\left|\frac{1}{n}\sum_{j=1}^{n} Y_{ij} - \exp\left\{\beta_1^0 + \frac{1}{2}(\sigma_v^2)^0 + U_i\right\}\phi(\boldsymbol{\beta}_{-1}^0)\right| > n^{-\rho}\right\} \\
= \;& O(n^{-C_2}), \qquad \text{(S2.33)}
\end{aligned}$$

which implies

$$\begin{aligned}
& P\left\{\left|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij} - \exp\left\{\beta_1^0 + \frac{1}{2}(\sigma_u^2)^0 + \frac{1}{2}(\sigma_v^2)^0\right\}\phi(\boldsymbol{\beta}_{-1}^0)\right| > n^{-\rho}\right\} \\
= \;& O(n^{-C_2}).
\end{aligned}$$

Combining results (S2.29) and (S2.31), similar to the derivation of (S2.33), we have

$$\begin{aligned} & P\left\{\max_{1\le j\le n}\left|\frac{1}{m}\sum_{i=1}^{m}Y_{ij}-\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+V_j\right\}\phi(\boldsymbol{\beta}_{-1}^0)\right|>m^{-\rho}\right\} \\ = \; & O(m^{-C_2}), \end{aligned} \tag{S2.34}$$

which implies

$$\begin{aligned} & P\left\{\left|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}-\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+\frac{1}{2}(\sigma_v^2)^0\right\}\phi(\boldsymbol{\beta}_{-1}^0)\right|>m^{-\rho}\right\} \\ = \; & O(m^{-C_2}). \end{aligned}$$

By the similar derivation of (S2.33) again, we can also obtain that

$$\begin{aligned} & P\Bigg\{\max_{1\le i\le m}\Bigg\|\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_{ij}Y_{ij} \\ & -\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+\frac{1}{2}(\sigma_v^2)^0\right\}\phi_1(\boldsymbol{\beta}_{-1}^0)\Bigg\|>n^{-\rho}\Bigg\}=O(n^{-C_2}), \end{aligned}$$

which implies

$$\begin{aligned} & P\Bigg\{\Bigg\|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij}Y_{ij} \\ & -\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+\frac{1}{2}(\sigma_v^2)^0\right\}\phi_1(\boldsymbol{\beta}_{-1}^0)\Bigg\|>n^{-\rho}\Bigg\}=O(n^{-C_2}), \end{aligned} \tag{S2.35}$$

and

$$\begin{aligned} & P\Bigg\{\max_{1\le i\le n}\Bigg\|\frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_{ij}Y_{ij} \\ & -\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+\frac{1}{2}(\sigma_v^2)^0\right\}\phi_1(\boldsymbol{\beta}_{-1}^0)\Bigg\|>m^{-\rho}\Bigg\}=O(m^{-C_2}), \end{aligned}$$

which implies

$$P\left\{\left\|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij}Y_{ij} - \exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0+\frac{1}{2}(\sigma_v^2)^0\right\}\phi_1(\boldsymbol{\beta}_{-1}^0)\right\| > m^{-\rho}\right\} = O(m^{-C_2}). \tag{S2.36}$$

In the following content, we denote $\beta_1^{r0} = \beta_1^0 + (\sigma_v^2)^0/2$ and $\beta_1^{c0} = \beta_1^0 + (\sigma_u^2)^0/2$.

Equating equations (S2.8), (S2.9), (S2.10), and (S2.11) to zero and solving, the estimates $\widehat{\underline{\mu}}_{u_i}, \widehat{\underline{\lambda}}_{u_i}, \widehat{\underline{\mu}}_{v_j}$, and $\widehat{\underline{\lambda}}_{v_j}$ can be obtained. And then plugging them into the equations, we have

$$\frac{1}{n}\sum_{j=1}^{n}Y_{ij} = \exp(\beta_1^r+\widehat{\underline{\mu}}_{u_i}+\widehat{\underline{\lambda}}_{u_i}/2)\frac{1}{n}\sum_{j=1}^{n}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) + \frac{\widehat{\underline{\mu}}_{u_i}}{n\sigma_u^2}, \tag{S2.37}$$

$$\frac{1}{n\widehat{\underline{\lambda}}_{u_i}} = \exp(\beta_1^r+\widehat{\underline{\mu}}_{u_i}+\widehat{\underline{\lambda}}_{u_i}/2)\frac{1}{n}\sum_{j=1}^{n}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) + \frac{1}{n\sigma_u^2} \tag{S2.38}$$

and

$$\frac{1}{m}\sum_{i=1}^{m}Y_{ij} = \exp(\beta_1^c+\widehat{\underline{\mu}}_{v_j}+\widehat{\underline{\lambda}}_{v_j}/2)\frac{1}{m}\sum_{i=1}^{m}\exp(\boldsymbol{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) + \frac{\widehat{\underline{\mu}}_{v_j}}{\sigma_v^2}, \tag{S2.39}$$

$$\frac{1}{m\widehat{\underline{\lambda}}_{v_j}} = \exp(\beta_1^c+\widehat{\underline{\mu}}_{v_j}+\widehat{\underline{\lambda}}_{v_j}/2)\frac{1}{m}\sum_{j=1}^{m}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) + \frac{1}{m\sigma_v^2}$$

where $(\boldsymbol{\beta}_{-1},\sigma_u^2,\sigma_v^2) = [-C_1,C_1]^{p-1}\times[-C_1,C_1]\times[C_1^{-1},C_1]$. According to (S2.28), (S2.29), (S2.33), (S2.34), and (S2.37), we obtain

$$\begin{aligned} & \exp(\beta_1^{r0}+U_i)\phi(\boldsymbol{\beta}_{-1}^0)+O_p(n^{-\rho}) \\ = \;& \exp\left(\beta_1^r+\widehat{\underline{\mu}}_{u_i}+\widehat{\underline{\lambda}}_{u_i}/2\right)\{\phi(\boldsymbol{\beta}_{-1})+O_p(n^{-\rho})\}+\frac{\widehat{\underline{\mu}}_{u_i}}{n\sigma_u^2}, \end{aligned} \tag{S2.40}$$

where the remainders $O_p(\cdot)$ are of the stated orders uniformly in $1 \le i \le m$ and in $(\beta_1^r, \boldsymbol{\beta}_{-1}, \sigma_u^2) \in [-2C_1, 2C_1] \times [-C_1, C_1]^{p-1} \times [C_1^{-1}, C_1]$. Since (S2.22) implies $\underline{\widehat{\lambda}}_{u_i} < \sigma_u^2$, the right-hand side of (S2.40) equals

$$
\begin{aligned}
& \{\underline{\widehat{\mu}}_{u_i}/(n\sigma_u^2)\} + \exp(\beta_1^r + \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2)\{\phi(\boldsymbol{\beta}_{-1}) + O_p(n^{-\rho})\} \\
= \; & \{\underline{\widehat{\mu}}_{u_i}/(n\sigma_u^2)\} + \exp(\mu_i + \omega_i)\{\phi(\boldsymbol{\beta}_{-1}) + O_p(n^{-\rho})\}, \qquad \text{(S2.41)}
\end{aligned}
$$

where $\beta_1^r - \sigma_u^2/2 \le \omega_i \le \beta_1^r + \sigma_u^2/2$. By the assumption (A5), $|\omega_i|$ is bounded uniformly in i. Together with uniformly bounded $|\omega_i|$, (S2.41), and (S2.40), we can derive that

$$
\begin{aligned}
& \exp(\beta_1^r + \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2)\{\phi(\boldsymbol{\beta}_{-1}) + o_p(1)\} \\
= \; & \exp(\beta_1^{r0} + U_i)\{\phi(\boldsymbol{\beta}_{-1}^0) + o_p(1)\} \qquad \text{(S2.42)}
\end{aligned}
$$

uniformly in $1 \le i \le m$ and in $(\beta_1^r, \boldsymbol{\beta}_{-1}, \sigma_u^2) \in [-2C_1, 2C_1] \times [-C_1, C_1]^{p-1} \times [C_1^{-1}, C_1]$. The detailed derivations of (S2.42) is similar to the process of getting the equation (4.13) in Hall et al. (2011).

By (S2.25) and (S2.38), we have

$$
(n\underline{\widehat{\lambda}}_{u_i})^{-1} = (n\sigma_u^2)^{-1} + \exp(\beta_1^r + \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2)\{\phi(\boldsymbol{\beta}_{-1}) + O_p(n^{-\rho})\} \qquad \text{(S2.43)}
$$

uniformly in $1 \le i \le m$ and in $(\beta_1^r, \boldsymbol{\beta}_{-1}, \sigma_u^2) \in [-2C_1, 2C_1] \times [-C_1, C_1] \times [C_1^{-1}, C_1]$. Combining (S2.42) and (S2.43) we derive that

$$
\underline{\widehat{\lambda}}_{u_i} = [\sigma_u^{-2} + n\exp(\beta_1^{r0} + U_i)\phi(\boldsymbol{\beta}_{-1}^0)\{1 + o_p(1)\}]^{-1},
$$

which implies that $0 < \underline{\widehat{\lambda}}_{u_i} < \sigma_u^2$. Using the fact $\max_{1\le i\le m}\exp(-U_i) = O(n^\epsilon)$, we have

$$\underline{\widehat{\lambda}}_{u_i} = \{n\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{r0}+U_i)\}^{-1} + o_p(n^{\epsilon-1}) = O_p(n^{\epsilon-1}). \quad \text{(S2.44)}$$

Combining (S2.42) and (S2.44) we have

$$\underline{\widehat{\mu}}_{u_i} = U_i + \beta_1^{r0} - \beta_1^r + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^0)}{\phi(\boldsymbol{\beta}_{-1})}\right\} + o_p(1). \quad \text{(S2.45)}$$

By (S2.29) and (S2.34), we have

$$\begin{aligned} & \exp(\beta_1^{c0}+V_j)\phi(\boldsymbol{\beta}_{-1}^0) + O_p(m^{-\rho}) \\ = \; & \exp\left(\beta_1^c + \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2\right)\{\phi(\boldsymbol{\beta}_{-1}) + O_p(n^{-\rho})\} + \frac{\underline{\widehat{\mu}}_{v_j}}{m\sigma_v^2} \end{aligned} \quad \text{(S2.46)}$$

uniformly in $1 \le j \le n$ and in $(\beta_1^c, \boldsymbol{\beta}_{-1}, \sigma_v^2) \in [-2C_1, 2C_1] \times [-C_1, C_1]^{p-1} \times [C_1^{-1}, C_1]$. The following properties can be derived by analogy (S2.42), (S2.44), and (S2.46),

$$\exp(\beta_1^c + \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2) = \exp(\beta_1^{c0}+V_j)\{\phi(\boldsymbol{\beta}_{-1}^0) + o_p(1)\},$$

$$\underline{\widehat{\lambda}}_{v_j} = \{m\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{c0}+V_j)\}^{-1} + o_p(m^{\epsilon-1}) = O_p(m^{\epsilon-1}), \quad \text{(S2.47)}$$

$$\underline{\widehat{\mu}}_{v_j} = V_j + \beta_1^{c0} - \beta_1^c + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^0)}{\phi(\boldsymbol{\beta}_{-1})}\right\} + o_p(1). \quad \text{(S2.48)}$$

By (S2.19), (S2.21), (S2.45), and (S2.48), we have

$$\beta_1^{r0} - \underline{\widehat{\beta}}_1^r + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^0)}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\right\} = o_p(1), \quad \text{(S2.49)}$$

$$\beta_1^{c0} - \underline{\widehat{\beta}}_1^c + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^0)}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\right\} = o_p(1). \quad \text{(S2.50)}$$

Combining (S2.5), (S2.35), (S2.36), (S2.40), and (S2.46) we have

$$\begin{aligned} & \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij}\left\{Y_{ij}-\exp\left(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\frac{1}{2}\lambda_{u_i}\right)\right\} \\ & +\frac{1}{nm}\sum_{j=1}^{n}\sum_{i=1}^{m}\mathbf{X}_{ij}\left\{Y_{ij}-\exp\left(\mathbf{X}_{ij}\boldsymbol{\beta}^{c}+\mu_{v_j}+\frac{1}{2}\lambda_{v_j}\right)\right\} \\ = \;& 2\phi(\boldsymbol{\beta}^{0}_{-1})\left\{\frac{\phi_1(\boldsymbol{\beta}^{0}_{-1})}{\phi(\boldsymbol{\beta}^{0}_{-1})}-\frac{\phi_1(\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1})}\right\}\exp\left\{\beta_1^0+\frac{(\sigma_u^2)^0}{2}+\frac{(\sigma_v^2)^0}{2}\right\}+o_p(1). \end{aligned} \tag{S2.51}$$

By the assumption (A9), (S2.51) implies the consistency of $\underline{\widehat{\boldsymbol{\beta}}}_{-1}$. Combining (S2.49) and (S2.50) with (S2.51), respectively, we obtain the consistency of $\underline{\widehat{\beta_1^r}}$ and $\underline{\widehat{\beta_1^c}}$.

Note that results (S2.12), (S2.44), and (S2.45) imply

$$\underline{\widehat{\sigma}}_u^2 = \frac{1}{m}\sum_{i=1}^{m}\left[U_i+\beta_1^{r0}-\beta_1^r+\log\left\{\frac{\phi(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}_{-1})}\right\}\right]^2+o_p(1),$$

by consistency of $\underline{\widehat{\beta_1^r}}$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1}$, and $m^{-1}\sum_{i=1}^m U_i^2=(\sigma_u^2)^0+O_p(m^{-1/2})$, further we have

$$\underline{\widehat{\sigma}}_u^2 = (\sigma_u^2)^0+o_p(1), \tag{S2.52}$$

which implies $\underline{\widehat{\sigma}}_u^2$ is consistent.

For the consistency of $\underline{\widehat{\sigma}}_v^2$, it can be derived from the results (S2.13), (S2.47), (S2.48), (S2.49), (S2.50), and $n^{-1}\sum_{j=1}^n V_j^2=(\sigma_v^2)^0+O_p(n^{-1/2})$ analogously:

$$\underline{\widehat{\sigma}}_v^2 = \frac{1}{n}\sum_{j=1}^{n}V_j^2+o_p(1)=(\sigma_v^2)^0+o_p(1). \tag{S2.53}$$

Due to $\beta_1^{r0} = \beta_1^0 + (\sigma_v^2)^0/2$ and $\beta_1^{c0} = \beta_1^0 + (\sigma_u^2)^0/2$, we have

$$\beta_1^0 \quad = \quad \frac{1}{2}\left\{\beta_1^{r0} - \frac{(\sigma_v^2)^0}{2} + \beta_0^{c0} - \frac{(\sigma_u^2)^0}{2}\right\} \tag{S2.54}$$

and can define the estimator of β_1^0 by

$$\widehat{\underline{\beta}}_1 \quad = \quad \frac{1}{2}\left(\widehat{\underline{\beta}_1^r} - \frac{\widehat{\underline{\sigma}}_v^2}{2} + \widehat{\underline{\beta}_1^c} - \frac{\widehat{\underline{\sigma}}_u^2}{2}\right). \tag{S2.55}$$

Together, equations (S2.49), (S2.50), (S2.52), (S2.53), (S2.54), and (S2.55) imply that

$$\widehat{\underline{\beta}}_1 \quad = \quad \beta_1^0 + o_p(1),$$

and β_1^0 can be estimated consistently.

S2.2 Convergence rate for variational estimators

Since (S2.24), (S2.26), and (S2.33) hold, we can define $\Delta_{i,1}, \Delta_{i,2}$, and $\Delta_{i,3}$ by

$$\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) \quad = \quad \phi(\boldsymbol{\beta}_{-1})\exp\{\Delta_{i,1}(\boldsymbol{\beta}_{-1})\}, \tag{S2.56}$$

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) \quad = \quad \phi_1(\boldsymbol{\beta}_{-1})\exp\{\Delta_{i,2}(\boldsymbol{\beta}_{-1})\}, \tag{S2.57}$$

$$\frac{1}{n\exp(\beta_1^{r0})}\left(\sum_{j=1}^{n} Y_{ij} - \frac{\mu_{u_i}}{\sigma_u^2}\right) \quad = \quad \phi(\boldsymbol{\beta}_{-1}^0)\exp\{U_i + \Delta_{i,3}(\boldsymbol{\beta}^r)\}, \tag{S2.58}$$

uniformly in $1 \leq i \leq m$; since (S2.25), (S2.27), and (S2.34) hold, we can define $\Delta'_{j,1}, \Delta'_{j,2}$, and $\Delta'_{j,3}$ by

$$\frac{1}{m}\sum_{i=1}^{m} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) = \phi(\boldsymbol{\beta}_{-1}) \exp\{\Delta'_{j,1}(\boldsymbol{\beta}_{-1})\}, \quad \text{(S2.59)}$$

$$\frac{1}{m}\sum_{i=1}^{m} \mathbf{X}_{ij} \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_{-1}) = \phi_1(\boldsymbol{\beta}_{-1}) \exp\{\Delta'_{j,2}(\boldsymbol{\beta}_{-1})\}, \quad \text{(S2.60)}$$

$$\frac{1}{m\exp(\beta_1^{c0})}\left(\sum_{i=1}^{m} Y_{ij} - \frac{\mu_{v_j}}{\sigma_v^2}\right) = \phi(\boldsymbol{\beta}_{-1}^{0}) \exp\{V_j + \Delta'_{j,3}(\boldsymbol{\beta}^{c})\}, \text{(S2.61)}$$

uniformly in j, and where $\Delta_{i,1}, \Delta_{i,2}, \Delta_{i,3}, \Delta'_{j,1}, \Delta'_{j,2}$, and $\Delta'_{j,3}$ satisfy, for all $C_1 > 0$ and each $\rho \in (0, 1/2)$,

$$\max_{1\leq i\leq n} \sup_{|\beta_1^r|,|\beta_2|,\ldots,|\beta_p|\leq C_1} |\Delta_{i,3}(\boldsymbol{\beta}^r)| = O_p(n^{-\rho}),$$

$$\max_{1\leq j\leq m} \sup_{|\beta_1^c|,|\beta_2|,\ldots,|\beta_p|\leq C_1} |\Delta'_{j,3}(\boldsymbol{\beta}^c)| = O_p(m^{-\rho}).$$

By Taylor series expansion of (S2.56), we have

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1})/\phi(\boldsymbol{\beta}_{-1}) \\ =\ & 1 + \Delta_{i,1}(\boldsymbol{\beta}_{-1}) + \frac{\Delta_{i,1}^2(\boldsymbol{\beta}_{-1})}{2} + O\{\Delta_{i,1}^3(\boldsymbol{\beta}_{-1})\}, \quad \text{(S2.62)}\end{aligned}$$

uniformly in $1 \leq i \leq m$, then squaring on both sides (S2.62), we have

$$\left\{\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1})/\phi(\boldsymbol{\beta}_{-1}) - 1\right\}^2 = \Delta_{i,1}^2(\boldsymbol{\beta}_{-1}) + O\{\Delta_{i,1}^3(\boldsymbol{\beta}_{-1})\},$$

which implies that for any $\epsilon > 0$,

$$\max_{1\leq i\leq m} \sup_{\boldsymbol{\beta}_{-1}:\|\boldsymbol{\beta}_{-1}\|\leq C_1} \Delta_{i,1}^2(\boldsymbol{\beta}_{-1}) = O_p(n^{\epsilon-1}). \quad \text{(S2.63)}$$

Combining (S2.63) and

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1})/\phi(\boldsymbol{\beta}_{-1})-1\\
=\ &\frac{1}{m}\sum_{i=1}^{m}\Delta_{i,1}(\boldsymbol{\beta}_{-1})+\frac{1}{m}\sum_{i=1}^{m}\frac{\Delta_{i,1}^{2}(\boldsymbol{\beta}_{-1})}{2}+O\{\Delta_{i,1}^{3}(\boldsymbol{\beta}_{-1})\}
\end{aligned}$$

we have

$$\sup_{\boldsymbol{\beta}_{-1}\in[-C_1,C_1]^{p-1}}\frac{1}{m}\left|\sum_{i=1}^{m}\Delta_{i,1}(\boldsymbol{\beta}_{-1})\right| = O_p\{(mn)^{\epsilon-1/2}+n^{\epsilon-1}\}. \tag{S2.64}$$

Similar to the derivation process of (S2.64), for any $\epsilon>0$, we have

$$\begin{aligned}
\sup_{\boldsymbol{\beta}_{-1}\in[-C_1,C_1]^{p-1}}\frac{1}{m}\left|\sum_{i=1}^{m}\Delta_{i,2}(\boldsymbol{\beta}_{-1})\right| &= O_p\{(mn)^{\epsilon-1/2}+n^{\epsilon-1}\}, \quad \text{(S2.65)}\\
\sup_{\boldsymbol{\beta}_{-1}\in[-C_1,C_1]^{p-1}}\frac{1}{n}\left|\sum_{j=1}^{n}\Delta'_{j,1}(\boldsymbol{\beta}_{-1})\right| &= O_p\{(mn)^{\epsilon-1/2}+m^{\epsilon-1}\},\\
\sup_{\boldsymbol{\beta}_{-1}\in[-C_1,C_1]^{p-1}}\frac{1}{n}\left|\sum_{j=1}^{n}\Delta'_{j,2}(\boldsymbol{\beta}_{-1})\right| &= O_p\{(mn)^{\epsilon-1/2}+m^{\epsilon-1}\}.
\end{aligned}$$

Using Taylor expansion of (S2.58) we have

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}Y_{ij}-\frac{\mu_{u_i}}{n\sigma_u^2} \qquad\qquad \text{(S2.66)}\\
=\ &\phi(\beta_1^0)\exp(U_i+\beta_1^{r0})\left\{1+\Delta_{i,3}(\boldsymbol{\beta}^r)+\frac{1}{2}\Delta_{i,3}^2(\boldsymbol{\beta}^r)+O_p(\Delta_{i,3}^3(\boldsymbol{\beta}^r))\right\}.
\end{aligned}$$

uniformly in $1\le i\le m$. By (S2.45), we have $\max_{1\le i\le m}|\mu_{u_i}|=O(n^{\epsilon})$ for each $\epsilon>0$. Combining this result and (S2.66) we have

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\phi^{-1}(\beta_1^0)\exp(-U_i-\beta_1^{r0})Y_{ij}-O(n^{\epsilon-1})-1\\
=\ &\Delta_{i,3}(\boldsymbol{\beta}^r)+\frac{1}{2}\Delta_{i,3}^2(\boldsymbol{\beta}^r)+O_p\{\Delta_{i,3}^3(\boldsymbol{\beta}^r)\}. \qquad\qquad \text{(S2.67)}
\end{aligned}$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$. Squaring on both sides (S2.67) we have

$$\begin{aligned}&\left\{\frac{1}{n}\sum_{j=1}^{n}\phi^{-1}(\boldsymbol{\beta}_{-1}^{0})\exp(-U_i-\beta_1^{r0})Y_{ij}-O_p(n^{\epsilon-1})-1\right\}^2\\ =\;& \Delta_{i,3}^2(\boldsymbol{\beta}^r)+O_p\{\Delta_{i,3}^3(\boldsymbol{\beta}^r)\} \qquad \text{(S2.68)}\end{aligned}$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$. Using the fact

$$E\left\{\frac{1}{n}\sum_{j=1}^{n}\phi^{-1}(\boldsymbol{\beta}_{-1}^{0})\exp(-U_i-\beta_1^{r0})Y_{ij}-1\right\}^2 = O_p(n^{-1}),$$

and (S2.68) we have

$$\max_{1\leq i\leq m}\sup_{|\beta_1^r|\leq 2C_1,\boldsymbol{\beta}_{-1}^0\in[-C_1,C_1]^{p-1}} O\{\Delta_{i,3}^2(\boldsymbol{\beta}^r)\} = O_p(n^{\epsilon-1}).$$

Combining the facts

$$E\left\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\phi^{-1}(\boldsymbol{\beta}_{-1}^{0})\exp(-U_i-\beta_1^{r0})Y_{ij}-1\right\}^2 = O\{(mn)^{-1}\}+O(n^{-1}),$$

and

$$\begin{aligned}&\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\phi^{-1}(\boldsymbol{\beta}_{-1}^{0})\exp(-U_i-\beta_1^{r0})Y_{ij}-O(n^{\epsilon-1})-1\\ =\;& \frac{1}{m}\sum_{i=1}^{m}\Delta_{i,3}(\boldsymbol{\beta}^r)+\frac{1}{m}\sum_{i=1}^{m}\frac{\Delta_{i,3}^2(\boldsymbol{\beta}^r)}{2}+O_p\{\Delta_{i,3}^3(\boldsymbol{\beta}^r)\},\end{aligned}$$

we have

$$\begin{aligned}&\sup_{|\beta_1^r|\leq 2C_1,\boldsymbol{\beta}_{-1}\in[-C_2,C_2]^{p-1}}\frac{1}{m}\left|\sum_{i=1}^{m}\Delta_{i,3}(\boldsymbol{\beta}^r)\right|\\ =\;& O_p\{n^{-1/2}+(mn)^{\epsilon-1/2}\}, \qquad \text{(S2.69)}\end{aligned}$$

for each $\epsilon > 0$. Similar to derivation of (S2.69), we can obtain

$$\begin{aligned} & \sup_{|\beta_1^r| \le 2C_1, \boldsymbol{\beta}_{-1} \in [-C_2, C_2]^{p-1}} \frac{1}{n} \left| \sum_{j=1}^{n} \Delta'_{j,3}(\beta_1^c, \beta_1) \right| \\ = \;& O_p\{m^{-1/2} + (mn)^{\epsilon - 1/2}\}, \qquad \text{(S2.70)} \end{aligned}$$

for each $\epsilon > 0$.

By the notation of (S2.56)-(S2.61), (S2.37) is equivalent to

$$\begin{aligned} & \exp\left\{ \beta_1^r + \underline{\widehat{\mu}}_{u_i} + \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \Delta_{i,1}(\boldsymbol{\beta}_{-1}) \right\} \phi(\boldsymbol{\beta}_{-1}) \\ = \;& \exp\{\beta_1^{r0} + U_i + \Delta_{i,3}(\boldsymbol{\beta}^r)\} \phi(\boldsymbol{\beta}_{-1}^0), \qquad \text{(S2.71)} \end{aligned}$$

uniformly in $1 \le i \le m$, and (S2.39) is equivalent to

$$\begin{aligned} & \exp\left\{ \beta_1^c + \underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + \Delta_{j,1}(\boldsymbol{\beta}_{-1}) \right\} \phi(\beta_1) \\ = \;& \exp\{\beta_1^{c0} + V_j + \Delta'_{j,3}(\boldsymbol{\beta}^c)\} \phi(\boldsymbol{\beta}_{-1}^0). \qquad \text{(S2.72)} \end{aligned}$$

uniformly in $1 \le j \le n$. Taking logarithms of both sides of (S2.71) and (S2.72), respectively, we have

$$\begin{aligned} & \beta_1^r - \beta_1^{r0} + \log\left\{ \frac{\phi(\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)} \right\} \\ = \;& U_i - \underline{\widehat{\mu}}_{u_i} - \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \Delta_{i,3}(\boldsymbol{\beta}^r) - \Delta_{i,1}(\boldsymbol{\beta}_{-\mathbf{1}}), \qquad \text{(S2.73)} \end{aligned}$$

uniformly in $1 \le i \le m$, and

$$\begin{aligned} & \beta_1^c - \beta_1^{c0} + \log\left\{ \frac{\phi(\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)} \right\} \\ = \;& V_j - \underline{\widehat{\mu}}_{v_j} - \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + \Delta'_{j,3}(\boldsymbol{\beta}^c) - \Delta'_{j,1}(\boldsymbol{\beta}_{-1}). \qquad \text{(S2.74)} \end{aligned}$$

uniformly in $1 \leq j \leq m$. Adding (S2.73) over i and adding (S2.74) over j, $\beta_1^r, \beta_1^c, \boldsymbol{\beta}_{-1}$ satisfy the following equations,

$$
\begin{aligned}
& \beta_1^r - \beta_1^{r0} + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)}\right\} \\
= \;& \frac{1}{m}\sum_{i=1}^m \left\{U_i - \underline{\widehat{\mu}}_{u_i} - \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \Delta_{i,3}(\boldsymbol{\beta}^r) - \Delta_{i,1}(\boldsymbol{\beta}_{-1})\right\}, \\
& \beta_1^c - \beta_1^{c0} + \log\left\{\frac{\phi(\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)}\right\} \\
= \;& \frac{1}{n}\sum_{j=1}^n \left\{V_j - \underline{\widehat{\mu}}_{v_j} - \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + \Delta'_{j,3}(\boldsymbol{\beta}^c) - \Delta'_{j,1}(\boldsymbol{\beta}_{-1})\right\}.
\end{aligned}
$$

By $m^{-1}\sum_{i=1}^m U_i = O_p(m^{-1/2})$, (S2.19), (S2.44), and the following property,

$$
\begin{aligned}
& \sup_{|\beta_1^r|\leq C_1, \boldsymbol{\beta}_{-1}\in[-C_1,C_1]} \frac{1}{m}\left|\sum_{i=1}^m \{\Delta_{i,3}(\boldsymbol{\beta}^0) - \Delta_{i,1}(\boldsymbol{\beta}_{-1})\}\right| \\
= \;& O_p\{n^{-1/2} + (mn)^{\epsilon-1/2}\}, \qquad \text{(S2.75)}
\end{aligned}
$$

which can be derived from (S2.64) and (S2.68), we have

$$
\underline{\widehat{\beta}}_1^r - \beta_1^{r0} + \log\left\{\frac{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)}\right\} = O_p(m^{-1/2} + n^{-1/2}). \qquad \text{(S2.76)}
$$

By $n^{-1}\sum_{j=1}^n V_j = O_p(n^{-1/2})$, (S2.21), (S2.47), and the following property,

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}\in[-C_1,C_1]^p} \frac{1}{n}\left|\sum_{i=1}^n \{\Delta'_{j,3}(\boldsymbol{\beta}^{r0}) - \Delta'_{j,1}(\boldsymbol{\beta}_{-1})\}\right| \\
= \;& O_p\{m^{-1/2} + (mn)^{\epsilon-1/2}\}, \qquad \text{(S2.77)}
\end{aligned}
$$

which can be derived from (S2.65) and (S2.69), we have

$$
\underline{\widehat{\beta}}_1^c - \beta_1^{c0} + \log\left\{\frac{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0)}\right\} = O_p(m^{-1/2} + n^{-1/2}). \qquad \text{(S2.78)}
$$

Define

$$\Delta_r = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\}, \qquad \text{(S2.79)}$$

and

$$\Delta_c = \frac{1}{nm}\sum_{j=1}^{n}\sum_{i=1}^{m}\mathbf{X}_{ij}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} + V_j)\}. \qquad \text{(S2.80)}$$

Since the variance of Δ_r is

$$\begin{aligned}
\mathrm{var}(\Delta_r) &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{ij,-1}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\}^2] \\
&\quad + \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l\neq j}^{n}E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{il,-1}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\} \\
&\quad \times\{Y_{il} - \exp(\mathbf{X}_{il}\boldsymbol{\beta}^{r0} + U_i)\}] \\
&\quad + \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m}E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{kj,-1}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\} \\
&\quad \times\{Y_{kj} - \exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_k)\}] \\
&= \frac{1}{m^2n^2}\mathrm{I} + \frac{1}{m^2n^2}\mathrm{II} + \frac{1}{m^2n^2}\mathrm{III},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{I} &= \sum_{i=1}^{m}\sum_{j=1}^{n}E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{ij,-1}\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\}^2] \\
&= mn\phi_2(\boldsymbol{\beta}_{-1}^0)\exp\{\beta_1^0 + (\sigma_u^2)^0/2 + (\sigma_v^2)^0/2\} \\
&\quad + mn\phi_2(2\boldsymbol{\beta}_{-1}^0)\exp\{2\beta_1^0 + 2(\sigma_u^2)^0 + (\sigma_v^2)^0\}[\exp\{(\sigma_v^2)^0\} - 1] \\
&= O(mn), \qquad \text{(S2.81)}
\end{aligned}$$

$$
\begin{aligned}
& \mathrm{II} && \text{(S2.82)}\\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l\neq j}^{n} E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{il,-1}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& \times\{Y_{il}-\exp(\mathbf{X}_{il}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\}] \\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l\neq j}^{n} \phi_1^2(\boldsymbol{\beta}_{-1}^0)\exp\{2(\sigma_u^2)^0\}E[\{\exp(\beta_1^0+V_j)-\exp(\beta_1^{r0})\} \\
& \times\{\exp(\beta_1^0+V_l)-\exp(\beta_1^{r0})\}] \\
= \; & 0, && \text{(S2.83)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathrm{III} \\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m} E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{kj,-1}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& \times\{Y_{kj}-\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_k)\}] \\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m} E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{kj,-1}E[\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& \times\{Y_{kj}-\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_k)\}|\mathbf{X}_{ij},\mathbf{X}_{kj},U_i,U_k,V_j]] \\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k\neq i}^{m} E[\mathbf{X}_{ij,-1}^{\mathrm{T}}\mathbf{X}_{kj,-1}\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0}+U_i+V_j)-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& \times\{\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{0}+U_k+V_j)-\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_k)\}] \\
= \; & m(m-1)n\phi_1^2(\boldsymbol{\beta}_{-1}^0)\exp\{2\beta_1^0+(\sigma_u^2)^0+(\sigma_v^2)^0\}[\exp\{(\sigma_v^2)^0\}-1] \\
= \; & O\{m(m-1)n\}. && \text{(S2.84)}
\end{aligned}
$$

Combining (S2.79), (S2.81), (S2.83), and (S2.84) we have

$$\Delta_r = O_p(n^{-1/2}).$$

The bound of Δ_c can be derived analogously,

$$\Delta_c = O_p(m^{-1/2}).$$

Setting (S2.5) equal to zero we have

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}^{\mathrm{T}}\left\{Y_{ij}-\exp\left(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}+\mu_{u_i}+\frac{1}{2}\lambda_{u_i}\right)\right\} && \text{(S2.85)}\\
&+\frac{1}{nm}\sum_{j=1}^{n}\sum_{i=1}^{m}\mathbf{X}_{ij,-1}^{\mathrm{T}}\left\{Y_{ij}-\exp\left(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c}+\mu_{v_j}+\frac{1}{2}\lambda_{v_j}\right)\right\}=0.
\end{aligned}$$

By (S2.57), (S2.60), (S2.79), and (S2.80), (S2.86) is equivalent to

$$\begin{aligned}
&\Delta_r+\exp(\beta_1^{r0})\phi_1(\boldsymbol{\beta}_{-1}^{0})\frac{1}{m}\sum_{i=1}^{m}\exp\{U_i+\Delta_{i,2}(\beta_1^{0})\}\\
&+\Delta_c+\exp(\beta_1^{c0})\phi_1(\boldsymbol{\beta}_{-1}^{0})\frac{1}{n}\sum_{j=1}^{n}\exp\{V_j+\Delta'_{j,2}(\boldsymbol{\beta}_{-1}^{0})\}\\
=\;&\exp(\beta_1^{r})\phi_1(\boldsymbol{\beta}_{-1})\frac{1}{m}\sum_{i=1}^{m}\exp\left\{\mu_{u_i}+\frac{1}{2}\lambda_{u_i}+\Delta_{i,2}(\boldsymbol{\beta}_{-1})\right\}\\
&+\exp(\beta_1^{c})\phi_1(\boldsymbol{\beta}_{-1})\frac{1}{n}\sum_{j=1}^{n}\exp\left\{\mu_{v_j}+\frac{1}{2}\lambda_{v_j}+\Delta'_{j,2}(\boldsymbol{\beta}_{-1})\right\} && \text{(S2.86)}
\end{aligned}$$

By equations (S2.71) and (S2.72), respectively,

$$\begin{aligned}
&\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i} && \text{(S2.87)}\\
=\;&\beta_1^{r0}-\underline{\widehat{\beta}}_1^{r}+\log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^{0})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\right\}+U_i+\Delta_{i,3}(\underline{\widehat{\boldsymbol{\beta}}}^{r})-\Delta_{i,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}),\\
&\underline{\widehat{\mu}}_{v_j}+\frac{1}{2}\underline{\widehat{\lambda}}_{v_j} && \text{(S2.88)}\\
=\;&\beta_1^{c0}-\underline{\widehat{\beta}}_1^{c}+\log\left\{\frac{\phi(\boldsymbol{\beta}_{-1}^{0})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\right\}+V_j+\Delta'_{j,3}(\underline{\widehat{\boldsymbol{\beta}}}^{c})-\Delta'_{j,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}),
\end{aligned}$$

we have the derivations of the two exponential terms on the right of (S2.86),

$$
\begin{aligned}
& \exp\left\{\underline{\widehat{\mu}}_{u_i} + \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \Delta_{i,2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\} && \text{(S2.89)}\\
= & \exp\left\{\beta_1^{r0} - \underline{\widehat{\beta_1^r}} + U_i + \Delta_{i,3}(\underline{\widehat{\boldsymbol{\beta}^r}}) - \Delta_{i,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) + \Delta_{i,2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\} \frac{\phi(\boldsymbol{\beta}^0_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})},\\
& \exp\left\{\underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + \Delta_{j,2}(\underline{\widehat{\boldsymbol{\beta}}}_{1})\right\} && \text{(S2.90)}\\
= & \exp\left\{\beta_1^{c0} - \underline{\widehat{\beta_1^c}} + V_j + \Delta'_{j,3}(\underline{\widehat{\boldsymbol{\beta}^c}}) - \Delta'_{j,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) + \Delta'_{j,2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\} \frac{\phi(\boldsymbol{\beta}^0_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}.
\end{aligned}
$$

Combining (S2.86), (S2.89), and (S2.90) we have

$$
\begin{aligned}
& \Delta_r \phi(\boldsymbol{\beta}^0_{-1})^{-1} + \exp(\beta_1^{r0}) \frac{\phi_1(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})} \frac{1}{m}\sum_{i=1}^m \exp\{U_i + \Delta_{i,2}(\boldsymbol{\beta}^0_{-1})\}\\
& + \Delta_c \phi(\boldsymbol{\beta}^0_{-1})^{-1} + \exp(\beta_1^{c0}) \frac{\phi_1(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})} \frac{1}{n}\sum_{j=1}^n \exp\{V_j + \Delta'_{j,2}(\boldsymbol{\beta}^0_{-1})\} && \text{(S2.91)}\\
= & \exp(\underline{\widehat{\beta_1^r}}) \frac{\phi_1(\boldsymbol{\beta}_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\\
& \times \frac{1}{m}\sum_{i=1}^m \exp\left\{\beta_1^{r0} - \underline{\widehat{\beta_1^r}} + U_i + \Delta_{i,3}(\underline{\widehat{\boldsymbol{\beta}^r}}) - \Delta_{i,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) + \Delta_{i,2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\}\\
& + \exp(\underline{\widehat{\beta_1^c}}) \frac{\phi_1(\boldsymbol{\beta}_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\\
& \times \frac{1}{n}\sum_{j=1}^n \exp\left\{\beta_1^{c0} - \underline{\widehat{\beta_1^c}} + V_j + \Delta'_{j,3}(\underline{\widehat{\boldsymbol{\beta}^c}}) - \Delta'_{j,1}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) + \Delta'_{j,2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\},
\end{aligned}
$$

By (S2.79), (S2.80), (S2.75), (S2.77), and (S2.91), we have

$$
\begin{aligned}
& \frac{\phi_1(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})} \left\{\exp(\beta_1^{r0}) \frac{1}{m}\sum_{i=1}^m \exp(U_i) + \exp(\beta_1^{c0}) \frac{1}{n}\sum_{j=1}^n \exp(V_j)\right\}\\
= & \frac{\phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})} \left\{\exp(\underline{\widehat{\beta_1^r}}) \frac{1}{m}\sum_{i=1}^m \exp(U_i) + \exp(\underline{\widehat{\beta_1^c}}) \frac{1}{n}\sum_{j=1}^n \exp(V_j)\right\}\\
& + O_p(n^{-1/2} + m^{-1/2}). && \text{(S2.92)}
\end{aligned}
$$

Substituting $\exp(\widehat{\underline{\beta}_1^r}) = \exp(\beta_1^r)\{1+(\widehat{\underline{\beta}_1^r}-\beta_1^{r0})+O_p(\widehat{\underline{\beta}_1^r}-\beta_1^{r0})\}$ and $\exp(\widehat{\underline{\beta}_1^c}) = \exp(\beta_1^c)\{1+(\widehat{\underline{\beta}_1^c}-\beta_1^{c0})+O_p(\widehat{\underline{\beta}_1^c}-\beta_1^{c0})\}$ into equation (S2.92), we have

$$\begin{aligned} & \frac{\phi_1(\boldsymbol{\beta}_{-1}^0)}{\phi(\boldsymbol{\beta}_{-1}^0)}\left\{\exp(\beta_1^{r0})\frac{1}{m}\sum_{i=1}^m \exp(U_i) + \exp(\beta_1^{c0})\frac{1}{n}\sum_{j=1}^n \exp(V_j)\right\} \\ = \;& \frac{\phi_1(\widehat{\underline{\boldsymbol{\beta}}}_{-1})}{\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})}\left\{\exp(\beta_1^{r0})\frac{1}{m}\sum_{i=1}^m \exp(U_i) + \exp(\beta_1^{c0})\frac{1}{n}\sum_{j=1}^n \exp(V_j)\right\} \\ & + O_p(n^{-1/2}+m^{-1/2}), \end{aligned}$$

which implies that

$$\phi_1(\boldsymbol{\beta}_{-1}^0)\phi(\boldsymbol{\beta}_{-1}^0)^{-1} = \phi_1(\widehat{\underline{\boldsymbol{\beta}}}_{-1})\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})^{-1} + O_p(m^{-1/2}+n^{-1/2}).$$

By the invertible mapping $\phi'(\cdot)/\phi(\cdot)$ and continuous mapping theorem, we have

$$\widehat{\underline{\boldsymbol{\beta}}}_{-1} = \boldsymbol{\beta}_{-1}^0 + O_p(m^{-1/2}+n^{-1/2}). \tag{S2.93}$$

Combining (S2.76), (S2.78), and (S2.93) we have

$$\begin{aligned} \widehat{\underline{\beta}_1^r} &= \beta_1^{r0} + O_p(m^{-1/2}+n^{-1/2}), & \text{(S2.94)} \\ \widehat{\underline{\beta}_1^c} &= \beta_1^{c0} + O_p(m^{-1/2}+n^{-1/2}). & \text{(S2.95)} \end{aligned}$$

Combining (S2.87), (S2.93), and (S2.94) we have

$$\widehat{\underline{\mu}}_{u_i} = U_i + \Delta_{i,3}(\widehat{\underline{\boldsymbol{\beta}}^r}) - \Delta_{i,1}(\widehat{\underline{\boldsymbol{\beta}}}_{-1}) + O_p(m^{-1/2}+n^{-1/2}), \tag{S2.96}$$

uniformly in $1 \le i \le m$. Together with (S2.75) and (S2.96) imply that

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^{m} \widehat{\underline{\mu}}_{u_i}^2 &= \frac{1}{m}\sum_{i=1}^{m} U_i^2 + O_p(m^{-1/2}+n^{-1/2}) \\ &= (\sigma_u^2)^0 + O_p(m^{-1/2}+n^{-1/2}). \end{aligned} \tag{S2.97}$$

And combining (S2.88), (S2.93), and (S2.95) we have

$$\widehat{\underline{\mu}}_{v_j} = V_j + \Delta'_{j,3}(\widehat{\underline{\boldsymbol{\beta}}^c}) - \Delta'_{j,1}(\widehat{\underline{\boldsymbol{\beta}}}_{-1}) + O_p(m^{-1/2}+n^{-1/2}), \tag{S2.98}$$

uniformly in $1 \le j \le m$. Equations (S2.77) and (S2.98) imply that

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} \widehat{\underline{\mu}}_{v_j}^2 &= \frac{1}{n}\sum_{j=1}^{n} V_j^2 + O_p(m^{-1/2}+n^{-1/2}) \\ &= (\sigma_v^2)^0 + O_p(m^{-1/2}+n^{-1/2}). \end{aligned} \tag{S2.99}$$

Combining (S2.12), (S2.44), and (S2.97) we have

$$\widehat{\underline{\sigma}}_u^2 = (\sigma_u^2)^0 + O_p(m^{-1/2}+n^{-1/2}), \tag{S2.100}$$

and combinig (S2.13), (S2.47), and (S2.99), we have

$$\widehat{\underline{\sigma}}_v^2 = (\sigma_v^2)^0 + O_p(m^{-1/2}+n^{-1/2}). \tag{S2.101}$$

Together, (S2.94), (S2.95), (S2.100), (S2.101), (S2.54), and (S2.55) give

$$\begin{aligned} \widehat{\underline{\beta}}_1 - \beta_1^0 &= \frac{1}{2}\left(\widehat{\underline{\beta}_1^r} - \beta_1^{r0} - \frac{\widehat{\underline{\sigma}}_v^2 - (\sigma_v^2)^0}{2} + \widehat{\underline{\beta}_1^c} - \beta_1^{c0} - \frac{\widehat{\underline{\sigma}}_u^2 - (\sigma_u^2)^0}{2}\right) \\ &= O_p(m^{-1/2}+n^{-1/2}). \end{aligned}$$

This completes the proof of consistency at rate $m^{-1/2}+n^{-1/2}$ for the Poisson regression model in Theorem 1.

S3 Proof of Theorem 2 for the Poisson regression model

The proof of Theorem 2 includes 7 steps:

Step1. formulas for estimators in §S3.1 ;

Step2. initial approximation to $\widehat{\underline{\beta}_1^r} - \beta_1^{r0}$, $\widehat{\underline{\beta}_1^c} - \beta_1^{c0}$, and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$ in §S3.2;

Step3. approximate formulae for $U_i, \widehat{\underline{\lambda}}_{u_i}, V_j$, and $\widehat{\underline{\lambda}}_{v_j}$ in §S3.3;

Step4. approximations to ζ_{u_i} and ζ_{v_j} in §S3.4;

Step5. final approximations to $\sigma_u^2 - (\sigma_u^2)^0$ and $\sigma_v^2 - (\sigma_v^2)^0$ in §S3.5;

Step6. approximations to ξ_{u_i}, η_{u_i}, ξ_{v_j}, and η_{v_j} in §S3.6;

Step7. final approximations to $\widehat{\underline{\beta}}_1 - \beta_1^0$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$ in §S3.7.

We first derive the initial approximation to $\widehat{\underline{\beta}_1^r} - \beta_1^{r0}$, $\widehat{\underline{\beta}_1^c} - \beta_1^{c0}$, and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$, and get the remainder terms of them , then control the remainder terms in §S3.2, §S3.3, §S3.5, and §S3.6. At last, §S3.4 and §S3.7 prove the asymptotic results of theorem by the results of remainder terms.

S3.1 Formulas for estimators

Based on the consistency results, we define ξ_{u_i}, η_{u_i}, and ζ_{u_i} by, respectively,

$$\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij,-1} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1} + V_j) \\ = \quad & \phi_1(\boldsymbol{\beta}^0_{-1}) \exp\{(\sigma_v^2)^0/2\} \exp(\xi_{u_i}), \end{aligned} \tag{S3.102}$$

$$\begin{aligned}
& \frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= \quad & \phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\exp(\eta_{u_i}), \qquad \text{(S3.103)}
\end{aligned}$$

and

$$\begin{aligned}
& \exp\left(\underline{\widehat{\beta^r}}_1+\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right)\frac{1}{n}\sum_{j=1}^{n}\left\{\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1})-\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\}+\frac{\underline{\widehat{\mu}}_{u_i}}{n\underline{\widehat{\sigma}}_u^2} \\
= \quad & \frac{1}{n}\sum_{j=1}^{n}\left\{Y_{ij}-\exp(\beta_1^{r0}+U_i)\phi(\boldsymbol{\beta}_{-1}^0)\right\} \\
& +\exp(\beta_1^{r0}+U_i)\phi(\boldsymbol{\beta}_{-1}^0)\left\{1-\exp(\zeta_{u_i})\right\}, \qquad \text{(S3.104)}
\end{aligned}$$

uniformly in $1 \leq i \leq m$; define ξ_{v_j}, η_{v_j}, and ζ_{v_j} by,

$$\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0+U_i) \\
= \quad & \phi_1(\boldsymbol{\beta}_{-1}^0)\exp\{(\sigma_u^2)^0/2\}\exp(\xi_{v_j}), \qquad \text{(S3.105)} \\
& \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= \quad & \phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\exp(\eta_{v_j}), \qquad \text{(S3.106)}
\end{aligned}$$

and

$$\begin{aligned}
& \exp\left(\underline{\widehat{\beta^c}}_1+\underline{\widehat{\mu}}_{v_j}+\frac{1}{2}\underline{\widehat{\lambda}}_{v_j}\right)\frac{1}{m}\sum_{i=1}^{m}\left\{\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1})-\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\}+\frac{\underline{\widehat{\mu}}_{v_j}}{m\underline{\widehat{\sigma}}_v^2} \\
= \quad & \frac{1}{m}\sum_{i=1}^{m}\left\{Y_{ij}-\exp(\beta_1^{c0}+V_j)\phi(\boldsymbol{\beta}_{-1}^0)\right\} \\
& +\exp(\beta_1^{c0}+V_j)\phi(\boldsymbol{\beta}_{-1}^0)\left\{1-\exp(\zeta_{v_j})\right\}, \qquad \text{(S3.107)}
\end{aligned}$$

uniformly in $1 \leq j \leq n$. By equations (S2.79) and (S2.80), (S2.5) is equivalent to

$$\begin{aligned}
& 2\Delta + \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0} + U_i + V_j) \\
& +\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{0} + U_i + V_j) \qquad \text{(S3.108)} \\
= \; & \exp(\underline{\widehat{\beta_1^r}})\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp\left(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1} + \underline{\widehat{\mu}}_{u_i} + \frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right) \\
& +\exp(\underline{\widehat{\beta_1^c}})\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp\left(\mathbf{X}_{ij,-1}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}}_{-1} + \underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j}\right),
\end{aligned}$$

where

$$\Delta = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\{Y_{ij} - \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + U_i + V_j)\},$$

using the notation in (S3.102)-(S3.107), we can write (S3.108) as

$$\begin{aligned}
& 2\Delta + \phi_1(\boldsymbol{\beta}_{-1}^{0})\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0} + U_i + \xi_{u_i}) \\
& +\phi_1(\boldsymbol{\beta}_{-1}^{0})\frac{1}{n}\sum_{j=1}^{n}\exp(\beta_1^{c0} + V_j + \xi_{v_j}) \\
= \; & \phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\frac{1}{m}\sum_{i=1}^{m}\exp\left(\underline{\widehat{\beta_1^r}} + \underline{\widehat{\mu}}_{u_i} + \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \eta_{u_i}\right) \\
& +\phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\frac{1}{n}\sum_{j=1}^{n}\exp\left(\underline{\widehat{\beta_1^c}} + \underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + \eta_{v_j}\right), \qquad \text{(S3.109)}
\end{aligned}$$

and write (S2.8) and (S2.10) as

$$\begin{aligned}
& \exp\left(\widehat{\underline{\beta}_1^r} + \underline{\widehat{\mu}}_{u_i} + \frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right)\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= & \exp(\beta_1^{r0} + U_i + \zeta_{u_i})\phi(\boldsymbol{\beta}^0_{-1}), \qquad \text{(S3.110)} \\
& \exp\left(\widehat{\underline{\beta}_1^c} + \underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j}\right)\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= & \exp(\beta_1^{c0} + V_j + \zeta_{v_j})\phi(\boldsymbol{\beta}^0_{-1}). \qquad \text{(S3.111)}
\end{aligned}$$

Substituting (S3.110) and (S3.111) into (S3.109) we obtain

$$\begin{aligned}
& \frac{2\Delta}{\phi(\boldsymbol{\beta}^0_{-1})} + \frac{\phi_1(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})}\frac{1}{m}\sum_{i=1}^m \exp(\beta_1^{r0} + U_i + \xi_{u_i}) \\
& + \frac{\phi_1(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})}\frac{1}{n}\sum_{j=1}^n \exp(\beta_1^{c0} + V_j + \xi_{v_j}) \\
= & \frac{\phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\frac{1}{m}\sum_{i=1}^m \exp\left(\beta_1^{r0} + U_i + \eta_{u_i} + \zeta_{u_i}\right) \\
& + \frac{\phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})}\frac{1}{n}\sum_{j=1}^n \exp\left(\beta_1^{c0} + V_j + \eta_{v_j} + \zeta_{v_j}\right), \qquad \text{(S3.112)}
\end{aligned}$$

S3.2 Initial approximation to $\widehat{\underline{\beta}_1^r} - \beta_1^{r0}$, $\widehat{\underline{\beta}_1^c} - \beta_1^{c0}$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$

To derive the approximation to $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$, write $\gamma(\boldsymbol{\beta}_{-1}) = \phi_1(\boldsymbol{\beta}_{-1})/\phi(\boldsymbol{\beta}_{-1})$, we have

$$\begin{aligned}
\gamma(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) &= \gamma(\boldsymbol{\beta}^0_{-1}) + \gamma'(\boldsymbol{\beta}^0_{-1})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) + O_p\{\|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\} \\
&= \gamma(\boldsymbol{\beta}^0_{-1}) + \{1 + O_p(m^{-1/2} + n^{-1/2})\}\gamma'(\boldsymbol{\beta}^0_{-1})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}),
\end{aligned}$$

where

$$\begin{aligned}\gamma'(\boldsymbol{\beta}^0_{-1}) &= \frac{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1^{\mathrm{T}}(\boldsymbol{\beta}^0_{-1})}{\phi(\boldsymbol{\beta}^0_{-1})^2},\\ \phi_2(\boldsymbol{\beta}_{-1}) &= E\{\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1})\} \qquad \text{(S3.113)}\end{aligned}$$

Therefore, by equation (S3.112),

$$\begin{aligned}&2\Delta/\phi(\boldsymbol{\beta}^0_{-1}) + \gamma(\boldsymbol{\beta}^0_{-1})\overline{S}_{U,\xi_u} + \gamma(\boldsymbol{\beta}^0_{-1})\overline{S}_{V,\xi_v}\\ =\;& [\gamma(\boldsymbol{\beta}^0_{-1}) + \{1 + O_p(m^{-1/2} + n^{-1/2})\}\gamma'(\boldsymbol{\beta}^0_{-1})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})](\overline{S}_{U,\eta_u,\zeta_u} + \overline{S}_{V,\eta_v,\zeta_v}),\end{aligned}$$

where

$$\begin{aligned}\overline{S}_{U,\xi_u} &= \frac{1}{m}\sum_{i=1}^{m}\exp\left(\beta_1^{r0} + U_i + \xi_{u_i}\right),\\ \overline{S}_{U,\eta_u,\zeta_u} &= \frac{1}{m}\sum_{i=1}^{m}\exp\left(\beta_1^{r0} + U_i + \eta_{u_i} + \zeta_{u_i}\right),\\ \overline{S}_{V,\xi_v} &= \frac{1}{n}\sum_{j=1}^{n}\exp\left(\beta_1^{c0} + V_j + \xi_{v_j}\right),\\ \overline{S}_{V,\eta_v,\zeta_v} &= \frac{1}{n}\sum_{j=1}^{n}\exp\left(\beta_1^{c0} + V_j + \eta_{v_j} + \zeta_{v_j}\right).\end{aligned}$$

And further, we have

$$\begin{aligned}&\gamma'(\beta_1^0)(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})(\overline{S}_{U,\eta_u,\zeta_u} + \overline{S}_{V,\eta_v,\zeta_v})\\ =\;& \frac{2\Delta}{\phi(\boldsymbol{\beta}^0_{-1})} + \gamma(\boldsymbol{\beta}^0_{-1})\{(\overline{S}_{U,\xi_u} - \overline{S}_{U,\eta_u,\zeta_u}) + (\overline{S}_{V,\xi_v} - \overline{S}_{V,\eta_v,\zeta_v})\}\\ &+O_p\{(m^{-1/2} + n^{-1/2})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})\}. \qquad \text{(S3.114)}\end{aligned}$$

Taking logarithms of both sides of (S3.110) and adding over i from 1 to m, then dividing by m we have

$$\begin{aligned}
& \log\{\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})/\phi(\boldsymbol{\beta}^0_{-1})\} \\
= \;& \beta_1^{r0} - \widehat{\underline{\beta}_1^r} + \frac{1}{m}\sum_{i=1}^m \left(U_i + \zeta_{u_i} - \widehat{\underline{\mu}}_{u_i} - \frac{1}{2}\widehat{\underline{\lambda}}_{u_i}\right),
\end{aligned} \tag{S3.115}$$

which implies that

$$\begin{aligned}
& \widehat{\underline{\beta}_1^r} - \beta_1^{r0} \\
= \;& -\gamma^{\mathrm{T}}(\boldsymbol{\beta}^0_{-1})(\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) \\
& + \frac{1}{m}\sum_{i=1}^m \left(U_i + \zeta_{u_i} - \widehat{\underline{\mu}}_{u_i} - \frac{1}{2}\widehat{\underline{\lambda}}_{u_i}\right) + O_p\{\|\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\}.
\end{aligned} \tag{S3.116}$$

Taking logarithms of both sides of (S3.111) and adding over j from 1 to n, then dividing by n we have

$$\log\{\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})/\phi(\boldsymbol{\beta}^0_{-1})\} = \beta_1^{c0} - \widehat{\underline{\beta}_1^c} + \frac{1}{n}\sum_{j=1}^n \left(V_j + \zeta_{v_j} - \widehat{\underline{\mu}}_{v_j} - \frac{1}{2}\widehat{\underline{\lambda}}_{v_j}\right),$$

which implies that

$$\begin{aligned}
\widehat{\underline{\beta}_1^c} - \beta_1^{c0} = \;& -\gamma^{\mathrm{T}}(\boldsymbol{\beta}^0_{-1})(\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) \\
& + \frac{1}{n}\sum_{j=1}^n \left(V_j + \zeta_{v_j} - \widehat{\underline{\mu}}_{v_j} - \frac{1}{2}\widehat{\underline{\lambda}}_{v_j}\right) + O_p\{\|\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\}).
\end{aligned}$$

S3.3 Approximate formulae for $U_i, \underline{\widehat{\lambda}}_{u_i}, V_j$ and $\underline{\widehat{\lambda}}_{v_j}$

Note that (S2.8) implies that

$$\begin{aligned}&\{1+O_p(n^{\epsilon-1/2})\}\phi(\beta_1^0)\exp(\beta_0^{r0}+U_i)\\ =\quad&\{1+O_p(n^{\epsilon-1/2})\}\phi(\beta_1)\exp\left(\beta_0^r+\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right)+\frac{\underline{\widehat{\mu}}_{u_i}}{n\underline{\widehat{\sigma}}_u^2},\end{aligned}$$

uniformly in $1\le i\le m$, for each $\epsilon>0$. Based on (S2.93) and (S2.94), we have

$$\begin{aligned}&\{1+O_p(n^{\epsilon-1/2})\}\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{r0}+U_i)\\ =\quad&\{1+O_p(n^{\epsilon-1/2})\}\phi(\boldsymbol{\beta}_{-1}^0)\exp\left(\beta_1^{r0}+\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right)+\frac{\underline{\widehat{\mu}}_{u_i}}{n\underline{\widehat{\sigma}}_u^2},\end{aligned}\tag{S3.117}$$

uniformly in $1\le i\le m$, for each $\epsilon>0$. Since $\max_{1\le i\le m}|U_i|=O_p\{(\log m)^{1/2}\}$, and (S2.22) implies that $0<\underline{\widehat{\lambda}}_{u_i}<\underline{\widehat{\sigma}}_u^2$, we have $\max_{1\le i\le m}|\underline{\widehat{\mu}}_{u_i}|=O_p\{(\log m)^{1/2}\}$. Therefore, by (S3.117),

$$\{1+O_p(n^{\epsilon-1/2})\}\exp(U_i)=\{1+O_p(n^{\epsilon-1/2})\}\exp\left(\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}\right),\tag{S3.118}$$

uniformly in $1\le i\le m$, for each $\epsilon>0$, taking logarithms, we have

$$U_i\quad=\quad\underline{\widehat{\mu}}_{u_i}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}+O_p(n^{\epsilon-1/2}),\tag{S3.119}$$

uniformly in $1\le i\le m$, for each $\epsilon>0$. Combining (S2.9), (S2.93), and (S2.94) we have

$$(n\underline{\widehat{\lambda}}_{u_i})^{-1}\quad=\quad\{1+O_p(n^{\epsilon-1/2})\}\phi(\boldsymbol{\beta}_{-1}^0)\exp\left(\underline{\widehat{\mu}}_i+\frac{1}{2}\underline{\widehat{\lambda}}_i+\beta_1^{r0}\right)+(n\underline{\widehat{\sigma}}_u^2)^{-1},$$

uniformly in $1 \le i \le m$, for $\epsilon > 0$. By (S3.118), we have

$$(n\underline{\widehat{\lambda}}_{u_i})^{-1} \;=\; \{1 + O_p(n^{\epsilon-1/2})\}\phi(\boldsymbol{\beta}^0_{-1}) \exp\left(U_i + \beta_1^{r0}\right) + (n\underline{\widehat{\sigma}}_u^2)^{-1},$$

uniformly in $1 \le i \le m$, for $\epsilon > 0$. Hence,

$$\underline{\widehat{\lambda}}_{u_i} \;=\; \{n\phi(\boldsymbol{\beta}^0_{-1}) \exp(U_i + \beta_1^{r0})\}^{-1} + O_p(n^{\epsilon-3/2}), \qquad \text{(S3.120)}$$

uniformly in $1 \le i \le m$, for $\epsilon > 0$. Based on (S2.9), (S2.93), and (S2.95), we can derive the formulae for V_j and $\underline{\widehat{\lambda}}_{v_j}$ analogously,

$$\begin{aligned} V_j &= \underline{\widehat{\mu}}_{v_j} + \frac{1}{2}\underline{\widehat{\lambda}}_{v_j} + O_p(m^{\epsilon-1/2}), \\ \underline{\widehat{\lambda}}_{v_j} &= \{m\phi(\boldsymbol{\beta}^0_{-1}) \exp(V_j + \beta_1^{c0})\}^{-1} + O_p(m^{\epsilon-3/2}), \end{aligned}$$

uniformly in $1 \le j \le n$, for $\epsilon > 0$.

S3.4 Approximations to ζ_{u_i} and ζ_{v_j}

Define the random variable δ_{u_i} from (S3.119) by

$$\underline{\widehat{\mu}}_{u_i} \;=\; U_i - \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} + \delta_{u_i}. \qquad \text{(S3.121)}$$

Denote $B^0_{u_i,k} = \sum_{j=1}^n \mathbf{X}^k_{ij,-1} \exp(\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^0 + V_j)$, $B^r_{u_i,k} = \sum_{j=1}^n \mathbf{X}^k_{ij,-1} \exp(\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^r)$ $B^{r0}_{u_i,k} = \sum_{j=1}^n \mathbf{X}^k_{ij,-1} \exp(\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^{r0})$ for $k = 0, 1$, and $\Delta_{u_i} = \sum_{j=1}^n Y_{ij} - B^0_{u_i,0}\exp(U_i)$ with zero mean and variance $n\phi(\boldsymbol{\beta}^0_{-1}) \exp\{\beta_1^0 + (\sigma_u^2)^0/2 +$

$(\sigma_v^2)^0/2\}$. We can write (S2.8) as

$$\begin{aligned} & \sum_{j=1}^{n} Y_{ij} - \sum_{j=1}^{n} \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{u_i} + \lambda_{u_i}/2) - \frac{\underline{\widehat{\mu}}_{u_i}}{\underline{\widehat{\sigma}}_u^2} \\ = & \ \Delta_{u_i} + B_{u_i,0}^{0}\exp(U_i) - B_{u_i,0}^{r}\exp(U_i+\delta_{u_i}) - \frac{\underline{\widehat{\mu}}_{u_i}}{\underline{\widehat{\sigma}}_u^2}, \quad \text{(S3.122)} \end{aligned}$$

uniformly in $1 \le i \le m$. By Taylor expansion of $\exp(U_i + \delta_{u_i})$, we have

$$\exp(U_i + \delta_{u_i}) = \exp(U_i)\left\{1 + \delta_{u_i} + \frac{1}{2}\delta_{u_i}^2 + O_p(n^{\epsilon-3/2})\right\},$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. By Taylor expression of $B_{u_i,0}^r = \sum_{j=1}^n \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r)$, we have

$$B_{u_i,0}^r = B_{u_i,0}^{r0} + \chi_{u_i} + O_p\{(m^{-1/2}+n^{-1/2})^3 n^{\epsilon+1}\},$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$, where

$$\begin{aligned} \chi_{u_i} = & \left\{\underline{\widehat{\beta}}_1^r - \beta_1^{r0} + \frac{1}{2}(\underline{\widehat{\beta}}_1^r - \beta_1^{r0})^2\right\} B_{u_i,0}^{r0} \\ & + \{\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0 + (\underline{\widehat{\beta}}_1 - \beta_1^0)(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}} B_{u_i,1}^{r0} \\ & + \frac{1}{2}(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} B_{u_i,2}^{r0} (\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0), \end{aligned}$$

where

$$B_{u_i,2}^{r0} = \sum_{j=1}^{n} \mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}} \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}). \quad \text{(S3.123)}$$

Therefore, (S3.122) implies that

$$
\begin{aligned}
& \sum_{j=1}^{n} Y_{ij} - B^{r}_{u_i,0} \exp(U_i + \delta_{u_i}) - \frac{\widehat{\underline{\mu}}_{u_i}}{\widehat{\underline{\sigma}}^2_u} \\
= \quad & \Delta_{u_i} + (B^{0}_{u_i,0} - B^{r0}_{u_i,0}) \exp(U_i) - B^{r0}_{u_i,0} \exp(U_i) \big\{ \delta_{u_i} + \frac{1}{2}\delta^2_{u_i} + O_p(n^{\epsilon-3/2}) \big\} \\
& -\chi_{u_i} \exp(U_i) \big\{ 1 + \delta_{u_i} + \frac{1}{2}\delta^2_{u_i} + O_p(n^{\epsilon-3/2}) \big\} - (U_i - \frac{1}{2}\widehat{\underline{\lambda}}_{u_i} + \delta_{u_i})/\widehat{\underline{\sigma}}^2_u \\
& +O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon+1}\},
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$, and further

$$
\begin{aligned}
& \delta_{u_i} + \frac{1}{2}\delta^2_{u_i} \frac{(B^{r0}_{u_i,0} + \chi_{u_i}) \exp(U_i)}{(B^{r0}_{u_i 0} + \chi_{u_i}) \exp(U_i) + 1/\widehat{\underline{\sigma}}^2_u} \\
= \quad & \frac{\Delta_{u_i} + (B^{0}_{u_i,0} - B^{r0}_{u_i,0}) \exp(U_i) - \chi_{u_i} \exp(U_i) - (U_i - \widehat{\underline{\lambda}}_{u_i}/2)/\widehat{\underline{\sigma}}^2_u}{(B^{r0}_{u_i,0} + \chi_{u_i}) \exp(U_i) + 1/\widehat{\underline{\sigma}}^2_u} \\
& +O_p(n^{\epsilon-3/2}) + O_p\{(m^{-1} + n^{-1})^3 n^{\epsilon}\},
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$, which implies that

$$
\begin{aligned}
\delta_{u_i} \quad = \quad & \frac{\Delta_{u_i} + (B^{0}_{u_i,0} - B^{r0}_{u_i,0}) \exp(U_i) - \chi_{u_i} \exp(U_i)}{(B^{r0}_{u_i,0} + \chi_{u_i}) \exp(U_i)} \\
& +O_p(n^{\epsilon-1}) + O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon}\} \\
= \quad & \{n \exp(\beta^{r0}_1) \phi(\boldsymbol{\beta}^0_{-1})\}^{-1} \Delta_{u_i} \exp(-U_i) \\
& +\{n \exp(\beta^{r0}_1) \phi(\boldsymbol{\beta}^0_{-1})\}^{-1} (B^{0}_{u_i,0} - B^{r0}_{u_i,0}) \\
& -(\widehat{\underline{\beta}^r_1} - \beta^{r0}_1) - \gamma^{\mathrm{T}}(\boldsymbol{\beta}^0_{-1})(\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) + O_p(n^{\epsilon-1}) \\
& +O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon}\}.
\end{aligned}
\tag{S3.124}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$. Hence, combining (S3.116), (S3.120), (S3.121), and (S3.124) we have

$$\begin{aligned}
\widehat{\underline{\mu}}_{u_i} &= U_i - \bar{U} + \bar{\zeta}_{u_i} + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
&\quad + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}\Delta_{u_i}\exp(-U_i) + O_p(n^{\epsilon-1}) \\
&\quad + O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon}\} \\
&= U_i - \bar{U} + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
&\quad + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}\Delta_{u_i}\exp(-U_i) \\
&\quad + O_p(n^{\epsilon-1}) + O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon}\}, \qquad \text{(S3.125)}
\end{aligned}$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$, where the second equality uses the property $\bar{\zeta}_{u_i} = O_p(n^{\epsilon-1})$ from the definition of ζ_{u_i}.

To derive the expansion of ζ_{u_i}, we substitute (S3.120) and (S3.125) into (S3.104),

$$\begin{aligned}
&\exp\left(\widehat{\underline{\mu}}_{u_i} + \frac{1}{2}\widehat{\underline{\lambda}}_{u_i} - U_i\right)\frac{1}{n}\sum_{j=1}^{n}\left\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\underline{\boldsymbol{\beta}}^{r}} + \exp(\widehat{\underline{\beta}_1^{r}})\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})\right\} \\
= \;& \frac{1}{n}\sum_{j=1}^{n}\left\{Y_{ij}\exp(-U_i) - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\right\} + \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\left\{1 - \exp(\zeta_{u_i})\right\} \\
&- \exp(-U_i)\frac{U_i}{n\widehat{\underline{\sigma}}_u^2} + O_p(n^{\epsilon-3/2}),
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$, where

$$
\begin{aligned}
& \exp\left(\widehat{\underline{\mu}}_{u_i} + \frac{1}{2}\widehat{\underline{\lambda}}_{u_i} - U_i\right) \\
= \; & 1 + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
& + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}\Delta_{u_i}\exp(-U_i) - \bar{U} \\
& + O_p(n^{\epsilon-1}) + O_p\{(m^{-1/2} + n^{-1/2})^3 n^{\epsilon}\},
\end{aligned}
$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$, therefore, we have

$$
\begin{aligned}
& [1 + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
& + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}\Delta_{u_i}\exp(-U_i) - \bar{U}] \\
& \times \frac{1}{n}\sum_{j=1}^n \left\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\underline{\boldsymbol{\beta}}}^r) - \exp(\widehat{\underline{\beta}_1^r})\phi(\widehat{\underline{\boldsymbol{\beta}}}_{-1})\right\} \\
= \; & \frac{1}{n}\sum_{j=1}^n \{Y_{ij}\exp(-U_i) - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\} \\
& - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\left(\zeta_{u_i} + \frac{1}{2}\zeta_{u_i}^2\right) \\
& - \exp(-U_i)\frac{U_i}{n\widehat{\underline{\sigma}}_u^2} + O_p(n^{\epsilon-3/2}),
\end{aligned}
$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Further we have

$$\begin{aligned}
& \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\left(\zeta_{u_i} + \frac{1}{2}\zeta_{u_i}^2\right) \qquad \text{(S3.126)} \\
= & -[1 + \{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
& +\{n\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\}^{-1}\Delta_{u_i}\exp(-U_i) - \bar{U}] \\
& \times\frac{1}{n}\sum_{j=1}^n\left\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}^r}) - \exp(\underline{\widehat{\beta}_1^r})\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\right\} \\
& +\frac{1}{n}\sum_{j=1}^n\left\{Y_{ij}\exp(-U_i) - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\right\} - \exp(-U_i)\frac{U_i}{n\underline{\widehat{\sigma}}_u^2} + O_p(n^{\epsilon-3/2}),
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Define

$$\begin{aligned}
D_k(b_1, \boldsymbol{b}_{-1}) &= \frac{1}{n}\sum_{j=1}^n\left\{\mathbf{X}_{ij}^k\exp(\mathbf{X}_{ij}^{\mathrm{T}}\mathbf{b}) - \exp(b_1)\phi_k(\boldsymbol{b}_{-1)}\right\} \\
&= O_p(n^{\epsilon-1/2}), k = 0, 1, \qquad \text{(S3.127)}
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Using the Taylor expression to $\exp(\mathbf{X}_{ij}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}^r}) - \exp(\underline{\widehat{\beta}_1^r})\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})$, we deduce that

$$\begin{aligned}
& \sum_{j=1}^n\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\underline{\widehat{\boldsymbol{\beta}}^r}) - \exp(\underline{\widehat{\beta}_1^r})\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})\} \\
= & n\{D_{u_i,0}(\boldsymbol{\beta}^{r0}) + (\underline{\widehat{\beta}_1^r} - \beta_1^{r0})D_{u_i,0}(\boldsymbol{\beta}^{r0}) + (\underline{\widehat{\beta}}_1 - \beta_1^0)D_{u_i,1}(\boldsymbol{\beta}^{r0})\} \\
& +O_p\{(m^{-1/2} + n^{-1/2})^2 n^{\epsilon+1/2}\}, \qquad \text{(S3.128)}
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Substituting (S3.127) and (S3.128) to (S3.126), and notice that,

$$\begin{aligned}
&\sum_{j=1}^{n}\{Y_{ij}\exp(-U_i) - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\} \\
= \;& \Delta_{u_i}\exp(-U_i) + \sum_{j=1}^{n}\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + V_j) - \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\},
\end{aligned}$$

we have

$$\begin{aligned}
&\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\left(\zeta_{u_i} + \frac{1}{2}\zeta_{u_i}^2\right) \\
= \;& \frac{1}{n}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
&+\frac{1}{n}\exp(-U_i)\left[\Delta_{u_i}\left\{1 - \exp(-\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)^{-1}D_0(\boldsymbol{\beta}^{r0})\right\} - U_i/\underline{\widehat{\sigma}}_u^2\right] \\
&-(\widehat{\underline{\beta}_1^r} - \beta_1^{r0})D_0(\boldsymbol{\beta}^{r0}) - D_1(\boldsymbol{\beta}^{r0})(\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} - \bar{U}D_0(\boldsymbol{\beta}^{r0}) \\
&+O_p\{(m^{-1/2} + n^{-1/2})^2 n^{\epsilon - 1/2}\}, \qquad \text{(S3.129)}
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$, and therefore,

$$\begin{aligned}
&\exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\zeta_{u_i} \\
= \;& \frac{1}{n}(B_{u_i,0}^0 - B_{u_i,0}^{r0}) \\
&+\frac{1}{n}\exp(-U_i)\left[\Delta_{u_i}\left\{1 - \exp(-\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)^{-1}D_0(\boldsymbol{\beta}^{r0})\right\} - U_i/\underline{\widehat{\sigma}}_u^2\right] \\
&-(\widehat{\underline{\beta}_1^r} - \beta_1^{r0})D_0(\boldsymbol{\beta}^{r0}) - (\widehat{\underline{\beta}}_1 - \beta_1^0)D_1(\boldsymbol{\beta}^{r0}) - \bar{U}D_0(\boldsymbol{\beta}^{r0}) \\
&-\frac{1}{2n^2}\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)(B_{u_i,0}^0 - B_{u_i,0}^{r0})^2 \\
&-\frac{1}{n^2}\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)(B_{u_i,0}^0 - B_{u_i,0}^{r0})\exp(-U_i)\Delta_{u_i} \\
&-\frac{1}{2n^2}\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)\exp(-2U_i)\Delta_{u_i}^2 + O_p(n^{\epsilon-3/2}) \\
&+O_p\{(m^{-1/2} + n^{-1/2})^2 n^{\epsilon - 1/2}\}, \qquad \text{(S3.130)}
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Equation (S3.130) implies that

$$\begin{aligned}
& \exp(\beta_1^{r0})\phi(\boldsymbol{\beta}_{-1}^0)\frac{1}{m}\sum_{i=1}^m U_i\zeta_{u_i} \\
= \; & \frac{1}{mn}\sum_{i=1}^m U_i(B_{u_i,0}^0 - B_{u_i,0}^{r0}) + \frac{1}{mn}\sum_{i=1}^m U_i\exp(-U_i)\Delta_{u_i} \\
& -\frac{1}{mn}\frac{1}{(\sigma_u^2)^0}\sum_{i=1}^m U_i^2\exp(-U_i) - \frac{\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)}{2mn^2}\sum_{i=1}^m U_i(B_{u_i,0}^0 - B_{u_i,0}^{r0})^2 \\
& -\frac{\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)}{mn^2}\sum_{i=1}^m U_i(B_{u_i,0}^0 - B_{u_i,0}^{r0})\exp(-U_i)\Delta_{u_i} \\
& -\frac{\exp(-\beta_1^{r0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)}{2mn^2}\sum_{i=1}^m U_i\exp(-2U_i)\Delta_{u_i}^2 \\
& +O_p(m^{-1/2}n^{\epsilon-1}) + O_p\{(m^{-1/2}+n^{-1/2})m^{-1/2}n^{\epsilon-1/2}\}. \qquad \text{(S3.131)}
\end{aligned}$$

for each $\epsilon > 0$. Since

$$\begin{aligned}
& E\left[\sum_{i=1}^m U_i\{B_{u_i,0}^0 - B_{u_i,0}^{r0}\}\right]^2 \\
= \; & mn(\sigma_u^2)^0\phi(2\boldsymbol{\beta}_{-1}^0)\exp\{2\beta_1^0 + (\sigma_v^2)^0\}\left[\exp\{(\sigma_v^2)^0\} - 1\right],
\end{aligned}$$

we have

$$\frac{1}{mn}\sum_{i=1}^m U_i(B_{u_i,0}^0 - B_{u_i,0}^{r0}) = O_p\{(mn)^{-1/2}\}. \qquad \text{(S3.132)}$$

Due to

$$\begin{aligned}
&\mathrm{var}\left\{ \sum_{i=1}^{m} U_i \exp(-U_i)\Delta_{u_i} \right\} \\
= \;& \sum_{i=1}^{m}\sum_{j=1}^{n} E\{U_i^2 \exp(-2U_i) \\
&\times E\left[\{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}^2 | \mathbf{X}_{ij}, U_i, V_j\right]\} \\
= \;& mn\phi(\boldsymbol{\beta}_{-1}^0)(\sigma_u^2)^0\{1 + (\sigma_u^2)^0\} \exp\left\{\beta_1^0 + \frac{1}{2}(\sigma_u^2)^0 + \frac{1}{2}(\sigma_v^2)^0\right\},
\end{aligned}$$

we have

$$\frac{1}{mn}\sum_{i=1}^{m} U_i \exp(-U_i)\Delta_{u_i} = O_p\{(mn)^{-1/2}\}. \tag{S3.133}$$

By (S3.132) and (S3.133), we have

$$\begin{aligned}
\frac{1}{mn^2}\sum_{i=1}^{m} U_i\{B_{u_i,0}^0 - B_{u_i,0}^{r0}\}^2 &= O_p(m^{-1/2}n^{\epsilon-1}) &\text{(S3.134)}\\
\frac{1}{mn^2}\sum_{i=1}^{m} U_i\{B_{u_i,0}^0 - B_{u_i,0}^{r0}\}\exp(-U_i)\Delta_{u_i} &= O_p(m^{-1/2}n^{\epsilon-1}) &\text{(S3.135)}
\end{aligned}$$

for each $\epsilon > 0$. Together with (S3.131), (S3.132), (S3.133), (S3.134), (S3.135), and the following facts

$$\begin{aligned}
E\{U_i^2 \exp(-U_i)\} &= \exp\{(\sigma_u^2)^0/2\}(\sigma_u^2)^0\{1 + (\sigma_u^2)^0\}, \\
E\left\{\sum_{i=1}^{m} U_i \exp(-2U_i)\Delta_{u_i}^2\right\} &= -mn(\sigma_u^2)^0\phi(\boldsymbol{\beta}_{-1}^0)\exp\{\beta_1^0 + (\sigma_u^2)^0/2 + (\sigma_v^2)^0/2\},
\end{aligned}$$

imply that

$$\frac{1}{m}\sum_{i=1}^{m} U_i\zeta_{u_i} = O_p\{(mn)^{-1/2} + n^{-1}\}.$$

Using (S3.130) and the following facts

$$\begin{aligned}\frac{1}{mn}\sum_{i=1}^{m}\exp(-U_i)\Delta_{u_i} &= O_p\{(mn)^{-1/2}\},\\ \frac{1}{mn}\sum_{i=1}^{m}\exp(-2U_i)\Delta_{u_i}^2 &= O_p(1),\end{aligned}$$

we obtain the following property,

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}\zeta_{u_i} &= \frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{r0})}\sum_{i=1}^{m}(B_{u_i,0}^0 - B_{u_i,0}^{r0})\\ &\quad +O_p\{(mn)^{-1/2}+n^{-1}\}. \qquad \text{(S3.136)}\end{aligned}$$

The following properties can be derived analogously:

$$\begin{aligned}&\exp(\beta_1^{c0})\phi(\boldsymbol{\beta}_{-1}^0)\zeta_{v_j}\\ = \;&\frac{1}{m}\{B_{v_j,0}^0 - B_{v_j,0}^{c0}\}\\ &+\frac{1}{m}\exp(-V_j)\left[\Delta_{v_j}\left\{1-\exp(-\beta_1^{c0})\phi(\boldsymbol{\beta}_{-1}^0)^{-1}D_0(\boldsymbol{\beta}^{c0})\right\} - V_j/\underline{\widehat{\sigma}}_v^2\right]\\ &-(\underline{\widehat{\beta}_1^c} - \beta_1^{c0})D_0(\boldsymbol{\beta}^{c0}) - D_1(\boldsymbol{\beta}^{c0})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0) - \bar{V}D_0(\boldsymbol{\beta}_0^{c0})\\ &-\frac{1}{2}\exp(-\beta_1^{c0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)m^{-2}\{B_{v_j,0}^0 - m\exp(\beta_1^{c0})\phi(\boldsymbol{\beta}_{-1}^0)\}^2\\ &-\exp(-\beta_1^{c0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)m^{-2}\{B_{v_j,0}^0 - m\exp(\beta_1^{c0})\phi(\boldsymbol{\beta}_{-1}^0)\}\exp(-V_j)\Delta_{v_j}\\ &-\frac{1}{2}\exp(-\beta_1^{c0})\phi^{-1}(\boldsymbol{\beta}_{-1}^0)m^{-2}\exp(-2V_j)\Delta_{v_j}^2\\ &+O_p(m^{\epsilon-3/2}) + O_p\{(m^{-1/2}+n^{-1/2})^2 m^{\epsilon-1/2}\},\end{aligned}$$

uniformly in $1 \le j \le n$, for each $\epsilon > 0$, and

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n} V_j \zeta_{v_j} &= O_p\{(mn)^{-1/2} + m^{-1}\}, \\
\frac{1}{n}\sum_{j=1}^{n} \zeta_{v_j} &= \{\phi(\boldsymbol{\beta}^0_{-1})\exp(\beta_1^{c0})\}^{-1}\frac{1}{mn}\sum_{j=1}^{n}(B^0_{v_j,0} - B^{c0}_{v_j,0}) \\
&\quad + O_p\{(mn)^{-1/2} + m^{-1}\}.
\end{aligned}$$

where $B^0_{v_j,0} = \sum_{i=1}^m \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i)$ and $B^{c0}_{v_j,0} = \sum_{i=1}^m \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0})$.

S3.5 Final approximations to $\sigma_u^2 - (\sigma_u^2)^0$ and $\sigma_v^2 - (\sigma_v^2)^0$

Equations (S3.110), (S3.115), and (S3.120) imply that

$$\begin{aligned}
\underline{\widehat{\mu}}_{u_i} &= U_i + \zeta_{u_i} - \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} - \log\{\phi(\underline{\widehat{\boldsymbol{\beta}}}_{-1})/\phi(\boldsymbol{\beta}^0_{-1})\} - (\underline{\widehat{\beta}_1^r} - \beta_1^{r0}) \qquad \text{(S3.137)} \\
&= U_i + \zeta_{u_i} - \frac{1}{2}\underline{\widehat{\lambda}}_{u_i} - \gamma(\boldsymbol{\beta}^0_{-1})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})^{\mathrm{T}} - (\underline{\widehat{\beta}_1^r} - \beta_1^{r0}) \\
&\quad + O_p\{(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})^{\mathrm{T}}(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})\} \\
&= U_i + \zeta_{u_i} - \{2n\phi(\boldsymbol{\beta}^0_{-1})\exp(U_i + \beta_1^{r0})\}^{-1} - (\bar{U} + \bar{\zeta}_u) \\
&\quad + \left[2n\phi(\boldsymbol{\beta}^0_{-1})\exp\left\{\beta_1^{r0} - \frac{1}{2}(\sigma^2)^0\right\}\right]^{-1} + O_p\{(m^{-1/2} + n^{-1/2})^2\}.
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Therefore, squaring both sides of (S3.137) and adding i from 1 to m, for $\epsilon > 0$,

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}\underline{\widehat{\mu}}^2_{u_i} &= \frac{1}{m}\sum_{i=1}^{m}(U_i + \zeta_{u_i} - \bar{U} - \bar{\zeta}_u)^2 \\
&\quad - \{mn\phi(\boldsymbol{\beta}^0_{-1})\exp(\beta_1^{r0})\}^{-1}\sum_{i=1}^{m}\exp(-U_i)(U_i + \zeta_{u_i} - \bar{U} - \bar{\zeta}_u) \\
&\quad + O_p\{(m^{-1/2} + n^{-1/2})^2\}. \qquad \text{(S3.138)}
\end{aligned}$$

Combining (S2.12), (S3.115), and (S3.138) we have

$$\begin{aligned}
\underline{\widehat{\sigma}}_u^2 &= \frac{1}{m}\sum_{i=1}^{m}(\underline{\widehat{\lambda}}_{u_i} + \underline{\widehat{\mu}}_{u_i}^2) \\
&= \frac{1}{m}\sum_{i=1}^{m}(U_i + \zeta_{u_i} - \bar{U} - \bar{\zeta}_u)^2 + O_p\{(m^{-1/2} + n^{-1/2})^2\} \quad \text{(S3.139)}
\end{aligned}$$

By (S3.136),

$$\begin{aligned}
\mathrm{var}(\bar{\zeta}_u) &= \frac{\{\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{r0})\}^{-2}}{m^2 n^2} \\
&\quad \times \sum_{i=1}^{m}\sum_{j=1}^{n} E\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\}^2 \\
&\quad + \frac{\{\phi(\boldsymbol{\beta}_{-1}^0)\exp(\beta_1^{r0})\}^{-2}}{m^2 n^2} \\
&\quad \times \sum_{i=1}^{m}\sum_{k\neq i}^{m}\sum_{j=1}^{n} E\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
&\quad \times\{\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
&= \frac{m-1}{mn}[\exp\{(\sigma_v^2)^0\} - 1] + O\{(mn)^{-1}\}, \qquad \text{(S3.140)}
\end{aligned}$$

which implies that

$$\bar{\zeta}_u = O_p(n^{-1/2}). \qquad \text{(S3.141)}$$

Moreover, by (S3.130),

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m} E\zeta_{u_i}^2 \\
&= \phi^{-2}(\boldsymbol{\beta}_{-1}^0)\exp(-2\beta_1^{r0})\frac{1}{mn^2}\sum_{i=1}^{m} E\{B_{u_i,0}^0 - B_{u_i,0}^{r0}\}^2 + O_p(m^{-1/2}n^{-1/2} + n^{-1}) \\
&= O_p(m^{-1/2}n^{-1/2} + n^{-1}).
\end{aligned}$$

Combining (S3.139), (S3.140), (S3.141), and (S3.136) we have

$$\begin{aligned}
& \widehat{\underline{\sigma}}_u^2 - (\sigma_u^2)^0 \\
= & \frac{1}{m}\sum_{i=1}^m (U_i + \zeta_{u_i} - \bar{U} - \bar{\zeta}_u)^2 - (\sigma_u^2)^0 \\
= & \frac{1}{m}\sum_{i=1}^m \{U_i^2 - (\sigma_u^2)^0\} + O_p\{(mn)^{-1/2} + m^{-1} + n^{-1}\}. \quad \text{(S3.142)}
\end{aligned}$$

By the assumption $\log m/\log n \to \delta$, $\delta \in (1/2, 2)$, (S3.142) implies the asymptotic normality result of $\widehat{\underline{\sigma}}_u^2$ in Theorem 2. The expression of $\widehat{\underline{\sigma}}_v^2$ can be derived analogously:

$$\widehat{\underline{\sigma}}_v^2 = \frac{1}{n}\sum_{j=1}^n (V_j + \zeta_{v_j} - \bar{V} - \bar{\zeta}_v)^2 + O_p\{(m^{-1/2} + n^{-1/2})^2\},$$

and

$$\widehat{\underline{\sigma}}_v^2 - (\sigma_v^2)^0 = \frac{1}{n}\sum_{j=1}^n \{V_j^2 - (\sigma_v^2)^0\} + O_p\{(m^{-1/2} + n^{-1/2})^2\}, \quad \text{(S3.143)}$$

which implies the asymptotic normality result of $\widehat{\underline{\sigma}}_v^2$ for the Poisson case in Theorem 2.

S3.6 Approximations to ξ_{u_i}, η_{u_i}, ξ_{v_j} and η_{v_j}

By (S3.102), we have

$$\begin{aligned}
& \frac{1}{n}\sum_{j=1}^n \mathbf{X}_{ij,-1} \exp\{\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1}^0 + V_j - (\sigma_v^2)^0/2\} \\
= & \phi_1(\boldsymbol{\beta}_{-1}^0)\{1 + \xi_{u_i} + \frac{1}{2}\xi_{u_i}^2 + O_p(|\xi_{u_i}|^3)\},
\end{aligned}$$

therefore,

$$\begin{aligned}
&\left(\xi_{u_i} + \frac{1}{2}\xi_{u_i}^2\right) \qquad \text{(S3.144)}\\
= \quad & \frac{\phi_1(\boldsymbol{\beta}_{-1}^0)^{\mathrm{T}}}{n\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\exp\{\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0 + V_j - (\sigma_v^2)^0/2\} - 1\\
&+O_p(|\xi_{u_i}|^3),
\end{aligned}$$

uniformly in $1 \leq i \leq m$. By (S3.130), we can derived that the term $\max_{1\leq i\leq m}|\xi_{u_i}| = O_p(n^{\epsilon-1/2})$. Squaring both sides of (S3.144), we have

$$\begin{aligned}
&\frac{1}{2}\xi_{u_i}^2\\
= \quad & \frac{1}{2\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2\exp\{(\sigma_v^2)^0\}}\\
&\times\left[\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} + V_j) - \|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2\exp\{(\sigma_v^2)^0/2\}\right]\\
&\times\left[\frac{1}{n}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} + V_j) - \|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2\exp\{(\sigma_v^2)^0/2\}\right]\\
&+O_p(n^{\epsilon-3/2}),
\end{aligned}$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$. Hence

$$
\begin{aligned}
& \xi_{u_i} \\
= \; & \frac{1}{\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_v^2)^0/2\}} \\
& \times \left[\frac{1}{n} \sum_{j=1}^n \mathbf{X}_{ij,-1}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}_{-1} + V_j) - \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_v^2)^0/2\} \right] \\
& - \frac{1}{2\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_v^2)^0\}} \\
& \times \left[\frac{1}{n} \sum_{j=1}^n \mathbf{X}_{ij,-1}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \exp(\mathbf{X}_{ij,-1} \boldsymbol{\beta}_{-1} + V_j) - \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_v^2)^0/2\} \right] \\
& \times \left[\frac{1}{n} \sum_{j=1}^n \mathbf{X}_{ij,-1}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \exp(\mathbf{X}_{ij,-1} \boldsymbol{\beta}_{-1} + V_j) - \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_v^2)^0/2\} \right] \\
& + O_p(n^{\epsilon - 3/2}),
\end{aligned}
$$

uniformly in $1 \le i \le m$, for each $\epsilon > 0$.

By (S3.103), we have

$$
\frac{1}{n} \sum_{j=1}^n \mathbf{X}_{ij,-1} \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \underline{\widehat{\boldsymbol{\beta}}}_{-1}) = \phi_1(\underline{\widehat{\boldsymbol{\beta}}}_{-1}) \left\{ 1 + \eta_{u_i} + \frac{1}{2}\eta_{u_i}^2 + O_p(|\eta_{u_i}|^3) \right\}.
$$

Based on the following fact,

$$
\begin{aligned}
& \frac{1}{n \phi_1(\underline{\widehat{\boldsymbol{\beta}}}^0_{-1}) \phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}} \sum_{j=1}^n \mathbf{X}_{ij,-1}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \underline{\widehat{\boldsymbol{\beta}}}_{-1}) - 1 \\
= \; & \frac{1}{n \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2} \sum_{j=1}^n \Big[\mathbf{X}_{ij,-1}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_1) \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}^0_{-1}) \\
& + \exp(\mathbf{X}_{ij,-1}^{\mathrm{T}} \boldsymbol{\beta}^0_{-1}) \mathbf{X}_{ij,-1}^{\mathrm{T}} (\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \\
& - \phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} \phi_1'(\boldsymbol{\beta}^0_{-1}) (\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) + O_p\{\|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\} \Big] \\
= \; & D_1(0, \boldsymbol{\beta}^0_{-1}) \phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} / \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 + O_p\{n^{\epsilon - 1/2} \|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|\},
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$, we obtain that

$$
\begin{aligned}
\eta_{u_i} + \frac{1}{2}\eta_{u_i}^2 &= D_1(0, \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} \phi_1(\boldsymbol{\beta}_{-1}^0)/\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2 \\
&\quad + O_p\{n^{\epsilon-1/2}\|\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0\| + |\eta_{u_i}|^3\}, \qquad \text{(S3.145)}
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$. Squaring both sides of (S3.145), we can deduce that

$$
\begin{aligned}
\eta_{u_i}^2 &= \{D_1(0, \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} \phi_1(\boldsymbol{\beta}_{-1}^0)\}^2/\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^4 \\
&\quad + O_p\{n^{\epsilon-1}\|\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0\| + |\eta_{u_i}|^3\}, \qquad \text{(S3.146)}
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$. Combining (S3.145) and (S3.146) we have

$$
\begin{aligned}
\eta_{u_i} &= D_1(0, \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} \phi_1(\boldsymbol{\beta}_{-1}^0)/\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2 \\
&\quad - \frac{1}{2}\{D_1(0, \boldsymbol{\beta}_{-1}^0)^{\mathrm{T}} \phi_1(\boldsymbol{\beta}_{-1}^0)\}^2/\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^4 \\
&\quad + O_p\{\|\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0\| n^{\epsilon-1/2} + |\eta_{u_i}|^3\}, \qquad \text{(S3.147)}
\end{aligned}
$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$.

For the approximations of ξ_{v_j} and η_{v_j}, we can derive them analogously,

$$
\begin{aligned}
\xi_{v_j} \quad = \quad & \frac{1}{\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_u^2)^0/2\}} \\
& \times \left[\frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}^0_{-1})\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1}+V_j) - \|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2\exp\{(\sigma_v^2)^0/2\}\right], \\
& -\frac{1}{2\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 \exp\{(\sigma_u^2)^0\}} \\
& \times \left[\frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}^{\mathrm{T}}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1}+V_j) - \phi_1(\boldsymbol{\beta}^0_{-1})\exp\{(\sigma_u^2)^0/2\}\right] \\
& \times \left[\frac{1}{n}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}^{\mathrm{T}}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1}+V_j) - \phi_1(\boldsymbol{\beta}^0_{-1})\exp\{(\sigma_u^2)^0/2\}\right] \\
& +O_p(m^{\epsilon-3/2}), \\
\eta_{v_j} \quad = \quad & D_1(0,\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})/\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2 - \frac{1}{2}\{D_1(0,\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})\}^2/\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^4 \\
& +O_p\{(m^{-1/2}+n^{-1/2})m^{\epsilon-1/2}\},
\end{aligned}
$$

uniformly in $1 \le j \le n$, for each $\epsilon > 0$.

S3.7 Final approximations to $\underline{\widehat{\beta}}_1 - \beta_1^0$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}$

For a true value $\boldsymbol{\beta}^0_{-1}$ and the bounded random variable $\mathbf{X}_{ij,-1}$'s, we have

$$
\begin{aligned}
& \operatorname{var}\left\{\frac{1}{m}\sum_{i=1}^{m} D_k(0,\boldsymbol{\beta}^0_{-1})\right\} \\
= \quad & \frac{1}{n^2m^2}\sum_{i=1}^{m}\sum_{j=1}^{n} E[\{\mathbf{X}_{ij,-1}^{k}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1}) - \phi_k(\boldsymbol{\beta}^0_{-1})\}^{\mathrm{T}} \\
& \times\{\mathbf{X}_{ij,-1}^{k}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1}) - \phi_k(\boldsymbol{\beta}^0_{-1})\}] \\
= \quad & \frac{1}{mn}\operatorname{var}\{\mathbf{X}_{-1}^{k}\exp(\mathbf{X}_{-1}^{\mathrm{T}}\boldsymbol{\beta}^0_{-1})\} = O(m^{-1}n^{-1}),
\end{aligned}
$$

furthermore,

$$\frac{1}{m}\sum_{i=1}^{m} D_k(0,\boldsymbol{\beta}_{-1}^0) = O_p(m^{-1/2}n^{-1/2}). \tag{S3.148}$$

By the equations (S3.130), (S3.147), and (S3.148), we have

$$\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i+\eta_{u_i}+\zeta_{u_i}) \\
= \;& \exp\left\{\beta_1^{r0}+\frac{1}{2}(\sigma_u^2)^0\right\}+O_p(n^{-1/2}).
\end{aligned} \tag{S3.149}$$

Analogously,

$$\begin{aligned}
& \frac{1}{n}\sum_{j=1}^{n}\exp(\beta_1^{c0}+V_j+\eta_{v_j}+\zeta_{v_j}) \\
= \;& \exp\left\{\beta_1^{c0}+\frac{1}{2}(\sigma_v^2)^0\right\}+O_p(m^{-1/2}).
\end{aligned} \tag{S3.150}$$

Combining (S3.130), (S3.144), (S3.147), and (S3.148) we have

$$\begin{aligned}
& \overline{S}_{U,\xi_u}-\overline{S}_{U,\eta_u,\zeta_u} \\
= \;& \frac{1}{m}\sum_{i=1}^{m}\left\{\exp(\beta_1^{r0}+U_i+\xi_{u_i})-\exp(\beta_1^{r0}+U_i+\eta_{u_i}+\zeta_{u_i})\right\} \\
= \;& \frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\left(\xi_{u_i}+\frac{1}{2}\xi_{u_i}^2-\eta_{u_i}-\frac{1}{2}\eta_{u_i}^2-\zeta_{u_i}-\frac{1}{2}\zeta_{u_i}^2-\eta_{u_i}\zeta_{u_i}\right) \\
& +O_p\left\{(m^{-1/2}+n^{-1/2})m^{-1/2}n^{\epsilon-1/2}+n^{\epsilon-3/2}\right\},
\end{aligned} \tag{S3.151}$$

for each $\epsilon > 0$. By (S3.144) and (S3.145), we have

$$\begin{aligned} & \xi_{u_i} + \frac{1}{2}\xi_{u_i}^2 - \eta_{u_i} - \frac{1}{2}\eta_{u_i}^2 \\ = \;& \frac{\phi_1(\boldsymbol{\beta}_{-1}^0)^{\mathrm{T}}}{\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2 \exp\{(\sigma_v^2)^0/2\}} \\ & \times \frac{1}{n}\sum_{j=1}^{n}[\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0 + V_j) - \mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0)\exp\{(\sigma_v^2)^0/2\}] \\ & + O_p\{n^{\epsilon-3/2}\}. \end{aligned}$$

Furthermore,

$$\begin{aligned} & \frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0} + U_i)(\xi_{u_i} + \frac{1}{2}\xi_{u_i}^2 - \eta_{u_i} - \frac{1}{2}\eta_{u_i}^2) \\ = \;& \frac{1}{mn\|\phi_1(\beta_1^0)\|^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j) \\ & - \mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i)\}\phi_1(\boldsymbol{\beta}_{-1}^0) + O_p(n^{\epsilon-3/2}). \end{aligned} \tag{S3.152}$$

By (S3.129), we have

$$\begin{aligned} & \frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0} + U_i)(\zeta_{u_i} + \frac{1}{2}\zeta_{u_i}^2) \\ = \;& \phi^{-1}(\boldsymbol{\beta}_{-1}^0)\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\exp(U_i)\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\ & + \phi^{-1}(\boldsymbol{\beta}_{-1}^0)\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}[\Delta_{u_i}\{1 - \phi^{-1}(\boldsymbol{\beta}_{-1}^0)\exp(-\beta_1^{r0})D_0(\boldsymbol{\beta}^{r0})\} - U_i/\underline{\widehat{\sigma}}_u^2] \\ & - \phi^{-1}(\boldsymbol{\beta}_{-1}^0)\frac{1}{m}\sum_{i=1}^{m}(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0)D_1(\boldsymbol{\beta}^{r0})\exp(\beta_1^{r0} + U_i) \\ & - \phi^{-1}(\boldsymbol{\beta}_{-1}^0)\frac{1}{m}\sum_{i=1}^{m}\bar{U}D_0(\boldsymbol{\beta}^{r0})\exp(\beta_1^{r0} + U_i) \\ & + O_p\{(m^{-1/2} + n^{-1/2})^2 n^{\epsilon-1/2}\}. \end{aligned} \tag{S3.153}$$

By (S3.130) and (S3.147), we can derive that

$$
\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\eta_{u_i}\zeta_{u_i} = \frac{1}{mn^2\phi(\boldsymbol{\beta}_{-1}^0)\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\\
&\times\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0)-\phi_1(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\\
&\times\sum_{j=1}^{n}\{Y_{ij}\exp(-U_i)-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\}+O_p(n^{\epsilon-3/2}).
\end{aligned}
$$

Since

$$
E\left\{\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\eta_{u_i}\zeta_{u_i}\right\} \quad = \quad 0,
$$

and

$$
\begin{aligned}
&E\left\{\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\eta_{u_i}\zeta_{u_i}\right\}^2 = \frac{1}{m^2n^4\phi(\boldsymbol{\beta}_{-1}^0)^2\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^4}\\
&\times E\Bigg(\sum_{i=1}^{m}\Bigg[\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0)-\phi_1(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\\
&\times\sum_{j=1}^{n}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\}\Bigg]^2\Bigg)+\frac{1}{m^2n^4\phi(\boldsymbol{\beta}_{-1}^0)^2\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^4}\\
&\times E\Bigg(\sum_{i=1}^{m}\sum_{k\neq i}^{m}\Bigg[\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}\exp(\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0)-\phi_1(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\\
&\times\sum_{j=1}^{n}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\}\Bigg]\\
&\times\Bigg[\sum_{j=1}^{n}\{\mathbf{X}_{kj,-1}\exp(\mathbf{X}_{kj,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^0)-\phi_1(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0)\\
&\times\sum_{j=1}^{n}\{Y_{kj}-\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_k)\}\Bigg]\Bigg)\\
=\;& O_p(m^{-1}n^{-2}+n^{-3}),
\end{aligned}
$$

we derive that

$$\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_1^{r0}+U_i)\eta_{u_i}\zeta_{u_i}=O_p(m^{-1/2}n^{-1}+n^{-3/2}). \qquad \text{(S3.154)}$$

Combining (S3.151), (S3.152), (S3.153), and (S3.154) we have

$$\begin{aligned}
& \overline{S}_{U,\xi_u}-\overline{S}_{U,\eta_u,\zeta_u} \\
= \; & \frac{1}{mn\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}-\gamma(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0) \\
& \times\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0+U_i+V_j)-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& +\frac{1}{mn\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\gamma(\boldsymbol{\beta}_{-1}^0)^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0) \\
& \times\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0+U_i+V_j)-\exp(\boldsymbol{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& -\frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0)}\sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
= \; & \frac{1}{mn\|\phi_1(\boldsymbol{\beta}_{-1}^0)\|^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}-\gamma(\boldsymbol{\beta}_{-1}^0)\}^{\mathrm{T}}\phi_1(\boldsymbol{\beta}_{-1}^0) \\
& \times\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0+U_i+V_j)-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}+U_i)\} \\
& -\frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0)}\sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}-\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0+U_i+V_j)\} \\
& +O_p(m^{\epsilon-3/2}+n^{\epsilon-3/2}), \qquad \text{(S3.155)}
\end{aligned}$$

for each $\epsilon > 0$. We can derive the following property analogously

$$\begin{aligned} & \overline{S}_{V,\xi_v} - \overline{S}_{V,\eta_v,\zeta_v} \\ = \; & \frac{1}{mn\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2} \sum_{i=1}^{m} \sum_{j=1}^{n} \{\mathbf{X}_{ij,-1} - \gamma(\boldsymbol{\beta}^0_{-1})\}^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \\ & \times\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} + V_j)\} \\ & - \frac{1}{mn\|\phi_1(\boldsymbol{\beta}^0_{-1})\|^2} \sum_{j=1}^{n} \gamma(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} \phi_1(\boldsymbol{\beta}^0_{-1}) \Delta_{v_j} \\ & + O_p(m^{\epsilon-3/2} + n^{\epsilon-3/2}), \end{aligned} \tag{S3.156}$$

for each $\epsilon > 0$, where $\Delta_{v_j} = \sum_{i=1}^{m} \{Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j)\}$.

Combining with (S3.114), (S3.149), (S3.150), (S3.155), and (S3.156) we have

$$\begin{aligned} & \gamma'(\boldsymbol{\beta}^0_{-1})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) 2\exp(\beta_1^0)\exp\{(\sigma_u^2)^0/2 + (\sigma_v^2)^0/2\} \\ = \; & T/\phi(\boldsymbol{\beta}^0_{-1}) + O_p\{(m^{-1/2} + n^{-1/2})(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1})\}, \end{aligned} \tag{S3.157}$$

where

$$\begin{aligned} T \; = \; & \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \{\mathbf{X}_{ij,-1} - \gamma(\boldsymbol{\beta}^0_{-1})\} \\ & \times\{2Y_{ij} - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} + U_i) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} + V_j)\}. \end{aligned}$$

T is asymptotically normally distributed with zero mean and variance

$$\begin{aligned}
&\mathrm{var}(T) \\
= \;& \frac{1}{mn} 4\{\phi_2(\boldsymbol{\beta}^0_{-1}) - \gamma(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\} \exp\{\beta^0_1 + (\sigma^2_u)^0/2 + (\sigma^2_v)^0/2\} \\
& + \frac{1}{mn}\{\phi_2(2\boldsymbol{\beta}^0_{-1}) - 2\gamma(\boldsymbol{\beta}^0_{-1})\phi_1(2\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} + \gamma(\boldsymbol{\beta}^0_{-1})\gamma(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\phi(2\boldsymbol{\beta}^0_{-1})\} \\
& \times \exp\{2\beta^0_1 + (\sigma^2_u)^0 + (\sigma^2_v)^0\} \\
& \times [4\exp\{(\sigma^2_u)^0 + (\sigma^2_v)^0\} - 3\exp\{(\sigma^2_u)^0\} - 3\exp\{(\sigma^2_v)^0\} + 2].
\end{aligned}$$

Therefore,

$$\sqrt{mn}T \rightarrow N(0, \boldsymbol{\Sigma}), \tag{S3.158}$$

in distribution, as $n \rightarrow \infty, m \rightarrow \infty$, where

$$\begin{aligned}
&\boldsymbol{\Sigma} \\
= \;& 4\{\phi_2(\boldsymbol{\beta}^0_{-1}) - \gamma(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\} \exp\{\beta^0_1 + (\sigma^2_u)^0/2 + (\sigma^2_v)^0/2\} \\
& + \{\phi_2(2\boldsymbol{\beta}^0_{-1}) - 2\gamma(\boldsymbol{\beta}^0_{-1})\phi_1(2\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} + \gamma(\boldsymbol{\beta}^0_{-1})\gamma(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\phi(2\boldsymbol{\beta}^0_{-1})\} \\
& \times \exp\{2\beta^0_1 + (\sigma^2_u)^0 + (\sigma^2_v)^0\} \\
& \times [4\exp\{(\sigma^2_u)^0 + (\sigma^2_v)^0\} - 3\exp\{(\sigma^2_u)^0\} - 3\exp\{(\sigma^2_v)^0\} + 2].
\end{aligned}$$

Combining (S3.157) and (S3.158) we have

$$\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1} \;=\; (mn)^{-1/2}\mathbf{Z}_3 + O_p(m^{-1}n^{-1/2} + m^{-1/2}n^{-1}),$$

where $\mathbf{Z}_3$ is a normal distribution with zero mean and variance

$$
\begin{aligned}
&\boldsymbol{\Sigma}_3 \\
= \quad & \frac{\phi^2(\boldsymbol{\beta}^0_{-1})}{\exp\{\beta^0_1 + (\sigma^2_u)^0/2 + (\sigma^2_v)^0/2\}} \{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}^{-1} \\
& \{\phi_2(\boldsymbol{\beta}^0_{-1}) - \gamma(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}\{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}^{-1} \\
& + \frac{1}{4}\phi^2(\boldsymbol{\beta}^0_{-1})\{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}^{-1} \\
& \times\{\phi_2(2\boldsymbol{\beta}^0_{-1}) - 2\gamma(\boldsymbol{\beta}^0_{-1})\phi_1(2\boldsymbol{\beta}^0_{-1})^{\mathrm{T}} + \gamma(\boldsymbol{\beta}^0_{-1})\gamma(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\phi(2\boldsymbol{\beta}^0_{-1})\} \\
& \times\{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}^{-1} \\
& \times[4\exp\{(\sigma^2_u)^0 + (\sigma^2_v)^0\} - 3\exp(\sigma^2_u)^0\} - 3\exp\{(\sigma^2_v)^0\} + 2].
\end{aligned}
$$

By (S3.116), (S3.117), (S3.142), and (S3.143), we have

$$
\begin{aligned}
&\widehat{\underline{\beta}}_1 - \beta^0_1 \\
= \quad & -\gamma(\beta^0_1)^{\mathrm{T}}(\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) + \frac{1}{2m}\sum_{i=1}^m (U_i + \zeta_{u_i}) \\
& + \frac{1}{2n}\sum_{j=1}^n \left(V_j + \zeta_{v_j}\right) - \frac{1}{4}\{\widehat{\underline{\sigma}}^2_u - (\sigma^2_u)^0 + \widehat{\underline{\sigma}}^2_v - (\sigma^2_v)^0\} \\
= \quad & \frac{1}{2}(\bar{U} + \bar{\zeta}_u + \bar{V} + \bar{\zeta}_v) - \frac{1}{4m}\sum_{i=1}^m\{U_i^2 - (\sigma^2_u)^0\} - \frac{1}{4n}\sum_{j=1}^n\{V_j^2 - (\sigma^2_v)^0\} \\
& + O_p\{(mn)^{-1/2} + m^{-1} + n^{-1}\} \\
:= \quad & S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + O_p\{(m^{-1/2} + n^{-1/2})^2\}, \qquad \text{(S3.159)}
\end{aligned}
$$

where

$$
\begin{aligned}
S_1 &= \frac{1}{2}\bar{\zeta}_u, \\
S_2 &= \frac{1}{2}\bar{\zeta}_v, \\
S_3 &= \frac{1}{2}\bar{U} = \frac{1}{2m}\sum_{i=1}^{m} U_i, \\
S_4 &= \frac{1}{2}\bar{V} = \frac{1}{2n}\sum_{j=1}^{n} V_j, \\
S_5 &= -\frac{1}{4m}\sum_{i=1}^{m}(U_i^2 - (\sigma_u^2)^0), \\
S_6 &= -\frac{1}{4n}\sum_{j=1}^{n}(V_j^2 - (\sigma_v^2)^0).
\end{aligned}
$$

By the requirements of multivariate central limit theorem, we need to check the boundedness of covariances of $(S_1, S_2, S_3, S_4, S_5, S_6)$. The diagonal elements can be calculated as follows,

$$
\begin{aligned}
\mathrm{var}(S_1) &= \frac{1}{4n}[\exp\{(\sigma_v^2)^0\} - 1] + O\{(mn)^{-1}\}, && \text{(S3.160)} \\
\mathrm{var}(S_2) &= \frac{1}{4m}[\exp\{(\sigma_u^2)^0\} - 1] + O\{(mn)^{-1}\} && \text{(S3.161)} \\
\mathrm{var}(S_3) &= \frac{1}{4m}(\sigma_u^2)^0, && \text{(S3.162)} \\
\mathrm{var}(S_4) &= \frac{1}{4n}(\sigma_v^2)^0, && \text{(S3.163)} \\
\mathrm{var}(S_5) &= \frac{1}{8m}\{(\sigma_u^2)^0\}^2, && \text{(S3.164)} \\
\mathrm{var}(S_6) &= \frac{1}{8n}\{(\sigma_v^2)^0\}^2. && \text{(S3.165)}
\end{aligned}
$$

The covariances can be computed as follows,

$$
\begin{aligned}
& \operatorname{cov}(S_1, S_2) \\
= \; & \frac{\exp(-\beta_1^{r0} - \beta_1^{c0})}{4m^2n^2\phi(\boldsymbol{\beta}_{-1}^0)^2} \sum_{i=1}^{m} \sum_{j=1}^{n} E[\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
& \times \{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0})\}] \\
& + \frac{\exp(-\beta_1^{r0} - \beta_1^{c0})}{4m^2n^2\phi(\boldsymbol{\beta}_{-1}^0)^2} \\
& \times \sum_{i=1}^{m} \sum_{k \neq i}^{m} \sum_{j=1}^{n} E[\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
& \times \{\exp(\mathbf{X}_{kj}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_k) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0})\}] \\
& + \frac{\exp(-\beta_1^{r0} - \beta_1^{c0})}{4m^2n^2\phi(\boldsymbol{\beta}_{-1}^0)^2} \\
& \times \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{l \neq j}^{n} E[\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
& \times \{\exp(\mathbf{X}_{il}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i) - \exp(\mathbf{X}_{il}^{\mathrm{T}}\boldsymbol{\beta}^{c0})\}] \\
= \; & O\{(mn)^{-1}\}, \qquad \text{(S3.166)}
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{cov}(S_1, S_4) \\
= \; & \frac{\exp(-\beta_1^{r0})}{4mn^2\phi(\boldsymbol{\beta}_{-1}^0)} \sum_{i=1}^{m} \sum_{j=1}^{n} E[V_j\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\}] \\
= \; & \frac{1}{4n}(\sigma_v^2)^0, \qquad \text{(S3.167)}
\end{aligned}
$$

$$
\begin{aligned}
& \operatorname{cov}(S_1, S_6) \\
= & -\frac{\exp(-\beta_1^{r0})}{8mn^2\phi(\boldsymbol{\beta}_{-1}^0)} \sum_{i=1}^{m}\sum_{j=1}^{n} E[\{\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + V_j) - \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0})\} \\
& \{V_j^2 - (\sigma_v^2)^0\}] \\
= & -\frac{\exp(\beta_1^0 - \beta_1^{r0})}{8n} E[\{V_j^2 - (\sigma_v^2)^0\}\exp(V_j)] \\
= & -\frac{1}{8n}\{(\sigma_v^2)^0\}^2, \qquad \text{(S3.168)}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{cov}(S_2, S_3) &= \frac{1}{4m}(\sigma_u^2)^0, & \text{(S3.169)} \\
\operatorname{cov}(S_2, S_5) &= -\frac{1}{8m}\{(\sigma_u^2)^0\}^2, & \text{(S3.170)}
\end{aligned}
$$

and others are zeros.

For simplicity, combining (S3.160)-(S3.170) we define two matrices by

$$
\boldsymbol{\Gamma}_1 = \frac{1}{8}\begin{pmatrix} 2[\exp\{(\sigma_u^2)^0\} - 1] & 2(\sigma_u^2)^0 & -\{(\sigma_u^2)^0\}^2 \\ 2(\sigma_u^2)^0 & 2(\sigma_u^2)^0 & 0 \\ -\{(\sigma_u^2)^0\}^2 & 0 & \{(\sigma_u^2)^0\}^2 \end{pmatrix} \qquad \text{(S3.171)}
$$

and

$$
\boldsymbol{\Gamma}_2 = \frac{1}{8}\begin{pmatrix} 2[\exp\{(\sigma_v^2)^0\} - 1] & 2(\sigma_v^2)^0 & -\{(\sigma_v^2)^0\}^2 \\ 2(\sigma_v^2)^0 & 2(\sigma_v^2)^0 & 0 \\ -\{(\sigma_v^2)^0\}^2 & 0 & \{(\sigma_v^2)^0\}^2 \end{pmatrix}. \qquad \text{(S3.172)}
$$

Combining (S3.172)-(S3.171) and following the multivariate central limit theorem we deduce that

$$
\begin{aligned}
&(\sqrt{m}(S_2, S_3, S_5), \sqrt{n}(S_1, S_4, S_6))^{\mathrm{T}} \\
\rightarrow \quad & N\left(\begin{pmatrix} \mathbf{0}_3 \\ \mathbf{0}_3 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma}_1 & 0 \\ 0 & \boldsymbol{\Gamma}_2 \end{pmatrix}\right), \qquad \text{(S3.173)}
\end{aligned}
$$

as $m, n \rightarrow \infty$, where $\mathbf{0}_3 = (0,0,0)^{\mathrm{T}}$. By results (S3.159), (S3.173), and delta method, we obtain

$$
\underline{\widehat{\beta}}_1 - \beta_1^0 \;\; = \;\; m^{-1/2} 1_3^{\mathrm{T}} Z_2 + n^{-1/2} 1_3^{\mathrm{T}} Z_1 + o_p(m^{-1/2} + n^{-1/2}),
$$

where Z_1 and Z_2 are normal distributions with mean zeros and variances $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$, respectively.

S4 Proof of Theorem 1 for Gamma regression model

The proof for the consistency and rates of convergence for the estimators $\underline{\widehat{\beta}}_1$, $\underline{\widehat{\boldsymbol{\beta}}}_{-1}$, $\underline{\widehat{\sigma}}_u^2$, and $\underline{\widehat{\sigma}}_v^2$ involves two main steps:

Step1. Establish the consistency of estimators $\beta_1^0, \boldsymbol{\beta}_{-1}^0, (\sigma_u^2)^0$, and $(\sigma_v^2)^0$ in §S4.1;

Step2. Extend the consistency results to a rate of convergence by controlling the remainder terms in §S4.2.

S4.1 Consistency

The variational lower bound of the CL for the Gamma regression model with crossed random effects is

$$
\begin{aligned}
& \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi}) \\
= & \sum_{i=1}^{m}\sum_{j=1}^{n}\left[\alpha Y_{ij}\left\{-\exp\left(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} - \mu_{u_i} + \frac{\lambda_{u_i}}{2}\right)\right\} - \alpha(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} + \mu_{u_i})\right] \\
& -\frac{m}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}) + \frac{1}{2}\log(\lambda_{u_i}) \\
& +\sum_{j=1}^{n}\sum_{i=1}^{m}\left[\alpha Y_{ij}\left\{-\exp\left(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} - \mu_{v_j} + \frac{\lambda_{v_j}}{2}\right)\right\} - \alpha(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} + \mu_{v_j})\right] \\
& -\frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{j=1}^{n}(\mu_{v_j}^2 + \lambda_{v_j}) + \frac{1}{2}\log\lambda_{v_j}.
\end{aligned}
$$

To derive the consistency of the estimates, we take derivatives of $\underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi})$ with respect to parameters,

$$
\begin{aligned}
\frac{\partial \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \beta_1^r} &= \sum_{i=1}^{m}\sum_{j=1}^{n}\alpha Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} - \mu_{u_i} + \lambda_{u_i}/2) \\
& -\alpha mn, \qquad \text{(S4.174)} \\
\frac{\partial \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \beta_1^c} &= \sum_{j=1}^{n}\sum_{i=1}^{m}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} - \mu_{v_j} + \lambda_{v_j}/2) - 1\} \\
& -\alpha mn, \qquad \text{(S4.175)}
\end{aligned}
$$

$$\begin{aligned}
&\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_{-1}} \qquad \text{(S4.176)}\\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} - \mu_{u_i} + \lambda_{u_i}/2) - 1\}\mathbf{X}_{ij,-1}\\
&\quad + \sum_{j=1}^{n}\sum_{i=1}^{m} \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} - \mu_{v_j} + \lambda_{v_j}/2) - 1\}\mathbf{X}_{ij,-1},
\end{aligned}$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \sigma_u^2} = -\frac{m}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}), \qquad \text{(S4.177)}$$

$$\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \sigma_v^2} = -\frac{m}{2\sigma_v^2} + \frac{1}{2\sigma_v^4}\sum_{i=1}^{m}(\mu_{v_j}^2 + \lambda_{v_j}), \qquad \text{(S4.178)}$$

for each μ_{u_i} and λ_{u_i},

$$\begin{aligned}
&\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \mu_{u_i}} \qquad \text{(S4.179)}\\
&= \sum_{j=1}^{n} \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} - \mu_{u_i} + \lambda_{u_i}/2) - 1\} - \frac{\mu_{u_i}}{\sigma_u^2},\\
&\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \lambda_{u_i}} \qquad \text{(S4.180)}\\
&= -\frac{1}{2}\sum_{j=1}^{n} \alpha Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r} - \mu_{u_i} + \lambda_{u_i}/2) - \frac{1}{2\sigma_u^2} + \frac{1}{2\lambda_{u_i}},
\end{aligned}$$

for each μ_{v_j} and λ_{v_j},

$$\begin{aligned}
&\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \mu_{v_j}} \qquad \text{(S4.181)}\\
&= \sum_{i=1}^{m} \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} - \mu_{v_j} + \lambda_{v_j}/2) - 1\} - \frac{\mu_{v_j}}{\sigma_u^2},\\
&\frac{\partial \underline{c\ell}(\mathbf{\Psi}^{rc}, \boldsymbol{\xi})}{\partial \lambda_{v_j}} \qquad \text{(S4.182)}\\
&= -\frac{1}{2}\sum_{j=1}^{n} \alpha Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c} - \mu_{v_j} + \lambda_{v_j}/2) - \frac{1}{2\sigma_v^2} + \frac{1}{2\lambda_{v_j}}.
\end{aligned}$$

Setting equations (S4.177) and (S4.178) to zeros and solving we have

$$\underline{\widehat{\sigma}}_u^2 = \frac{1}{m}\sum_{i=1}^{m}(\underline{\widehat{\mu}}_{u_i}^2 + \underline{\widehat{\lambda}}_{u_i}), \tag{S4.183}$$

$$\underline{\widehat{\sigma}}_v^2 = \frac{1}{n}\sum_{j=1}^{n}(\underline{\widehat{\mu}}_{v_j}^2 + \underline{\widehat{\lambda}}_{v_j}). \tag{S4.184}$$

Setting (S4.174), (S4.175), (S4.179), and (S4.181) to zeros we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - \mu_{u_i} + \lambda_{u_i}/2) - 1\} = 0, \tag{S4.185}$$

$$\sum_{j=1}^{n}\sum_{i=1}^{m}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c - \mu_{v_j} + \lambda_{v_j}/2) - 1\} = 0, \tag{S4.186}$$

$$\sum_{j=1}^{n}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - \mu_{u_i} + \lambda_{u_i}/2) - 1\} - \frac{\mu_{u_i}}{\sigma_u^2} = 0, \tag{S4.187}$$

$$\sum_{i=1}^{m}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c - \mu_{u_i} + \lambda_{v_j}/2) - 1\} - \frac{\mu_{v_j}}{\sigma_v^2} = 0. \tag{S4.188}$$

for $1 \le j \le n$. By (S4.187), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - \mu_{u_i} + \lambda_{u_i}/2) - 1\} = \sum_{i=1}^{m}\frac{\mu_{u_i}}{\sigma_u^2}. \tag{S4.189}$$

Combining (S4.189) and (S4.185) we have

$$\sum_{i=1}^{m}\underline{\widehat{\mu}}_{u_i} = 0. \tag{S4.190}$$

By (S4.188), we have

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c - \mu_{u_i} + \lambda_{v_j}/2) - 1\} = \sum_{j=1}^{n}\frac{\mu_{v_j}}{\sigma_v^2}. \tag{S4.191}$$

Combining (S4.191) and (S4.186) we have

$$\sum_{j=1}^{n}\underline{\widehat{\mu}}_{v_j} = 0. \tag{S4.192}$$

Setting equations (S4.180) and (S4.184) to zeros we have $\underline{\widehat{\lambda}}_{u_i}$ and $\underline{\widehat{\lambda}}_{v_j}$ satisfy

$$\underline{\widehat{\lambda}}_{u_i} = \left\{\sigma_u^{-2} + \sum_{j=1}^{n} \alpha Y_{ij} \exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2)\right\}^{-1}, \quad \text{(S4.193)}$$

and

$$\underline{\widehat{\lambda}}_{v_j} = \left\{\sigma_v^{-2} + \sum_{i=1}^{m} \alpha Y_{ij} \exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c - \underline{\mu}_{v_j} + \underline{\lambda}_{v_j}/2)\right\}^{-1},$$

from which it is obvious that $0 < \underline{\widehat{\lambda}}_{u_i} < \sigma_u^2$ and $0 < \underline{\widehat{\lambda}}_{v_j} < \sigma_v^2$.

Since conditional on X_{ij}, U_i, and V_j, Y_{ij}'s are independent Gamma regression with mean $\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j), 1 \le i \le m, 1 \le j \le n$, we have

$$E\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - U_i - V_j)\} = \exp(\beta_1^0 - \beta_1)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}). \quad \text{(S4.194)}$$

By (S4.194), the Markov's inequality and $m/n = O(1)$, for all C_1, C_2, and $\rho \in (0, 1/2)$, we have

$$\begin{aligned} &P\left[\max_{1\le i\le m}\left|\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - U_i - V_j\}\right.\right. \\ &\left.\left.- \exp(\beta_1^0 - \beta_1)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\}\right| > n^{-\rho}\right] = O(n^{-C_2}). \quad \text{(S4.195)} \end{aligned}$$

Combining (S2.28), (S2.30), and (S4.195), for all $C_1, C_2 > 0$ and $\rho \in (1, 1/2)$, we have

$$\begin{aligned} &P\left\{\max_{1\le i\le m}\left|\frac{1}{n}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - U_i)\right.\right. \quad \text{(S4.196)} \\ &\left.\left.- \exp\left\{\beta_1^0 + \frac{1}{2}(\sigma_u^2)^0 - \beta_1^r\right\}\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\right| > n^{-\rho}\right\} = O(n^{-C_2}), \end{aligned}$$

which implies

$$P\Big\{\Big|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r-U_i)$$
$$-\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0-\beta_1^r\right\}\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})\Big|>n^{-\rho}\Big\}=O(n^{-C_2}),$$

and

$$P\Big\{\Big|\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r-U_i) \tag{S4.197}$$
$$-\exp\left\{\beta_1^0+\frac{1}{2}(\sigma_u^2)^0-\beta_1^r\right\}\phi_1(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})\Big|>n^{-\rho}\Big\}=O(n^{-C_2}).$$

Equating equations (S4.179)-(S4.183) to zeros and solving, the estimates $\widehat{\underline{\mu}}_{u_i}$, $\widehat{\underline{\lambda}}_{u_i}$, $\widehat{\underline{\mu}}_{v_j}$, and $\widehat{\underline{\lambda}}_{v_j}$, can be obtained. Then plugging them into the equations, we have

$$\frac{1}{n}\sum_{j=1}^{n}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r-\widehat{\underline{\mu}}_{u_i}+\widehat{\underline{\lambda}}_{u_i}/2)-1\}=\frac{\widehat{\underline{\mu}}_{u_i}}{n\sigma_u^2}, \tag{S4.198}$$

and

$$-\frac{1}{n}\sum_{j=1}^{n}\alpha Y_{ij}\exp(\widehat{\underline{\lambda}}_{u_i}/2-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r-\widehat{\underline{\mu}}_{u_i})=\frac{1}{n}(\frac{1}{\sigma_u^2}-\frac{1}{\lambda_{u_i}}), \tag{S4.199}$$

uniformly in $1\le i\le m$, and

$$\frac{1}{m}\sum_{i=1}^{m}\alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c-\widehat{\underline{\mu}}_{v_j}+\widehat{\underline{\lambda}}_{v_j}/2)-1\}=\frac{\widehat{\underline{\mu}}_{v_j}}{m\sigma_v^2}, \tag{S4.200}$$

and

$$-\frac{1}{m}\sum_{i=1}^{m}\alpha Y_{ij}\exp(\widehat{\underline{\lambda}}_{v_j}/2-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c-\widehat{\underline{\mu}}_{v_j})=\frac{1}{m}(\frac{1}{\sigma_v^2}-\frac{1}{\lambda_{v_j}}), \tag{S4.201}$$

uniformly in $1 \le j \le n$, where $(\beta_1^r, \beta_1^c, \boldsymbol{\beta}_{-1}, \sigma_u^2, \sigma_v^2) = [-2C_1, 2C_1]\times[-2C_1, C_1]\times [C_1^{-1}, C_1]^{p-1} \times [C_1^{-1}, C_1]$. By (S4.196), (S4.198) and (S4.199) can be written as

$$\begin{aligned}
&\exp(\beta_1^{r0} - \beta_1^r + U_i - \underline{\widehat{\mu}}_i + \underline{\widehat{\lambda}}_{u_i}/2)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) + O_p(n^{-\rho}) - 1 \\
= \;& \frac{\underline{\widehat{\mu}}_{u_i}}{n\alpha\sigma_u^2}, \qquad \text{(S4.202)}
\end{aligned}$$

and

$$\begin{aligned}
&\exp(\beta_1^{r0} - \beta_1^r + U_i - \underline{\widehat{\mu}}_i + \underline{\widehat{\lambda}}_{u_i}/2)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) + O_p(n^{-\rho}) \\
= \;& \frac{1}{n\alpha\underline{\widehat{\lambda}}_{u_i}} - \frac{1}{n\alpha\sigma_u^2} \qquad \text{(S4.203)}
\end{aligned}$$

where the remainders $O_p(\cdot)$ are of the stated orders uniformly in $1 \le i \le m$ and in $(\beta_1^r, \boldsymbol{\beta}_{-1}, \sigma_u^2) \in [-2C_1, 2C_1] \times [-C_1, C_1]^{p-1} \times [C_1^{-1}, C_1]$. By (S4.202), we have

$$\begin{aligned}
&\exp(\beta_1^{r0} + U_i)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) \\
= \;& \{1 + \mu_{u_i}/(n\alpha\sigma_u^2) + O_p(n^{-\rho})\} \exp(\beta_1^r + \mu_{u_i} - \lambda_{u_i}/2) \quad \text{(S4.204)}
\end{aligned}$$

By (S4.193), $\lambda_{u_i} < \sigma_u^2$. Therefore, the term $\exp(\beta_1^r + \mu_{u_i} - \lambda_{u_i}/2)$ in the right-hand side of (S4.204) equals $\exp(\mu_{u_i} + \omega_i)$, where $\beta_1^r - \sigma_u^2/2 \le \omega_i \le \beta_1^r + \sigma_u^2/2$, Since σ_u^2 and β_1^r are in compact sets and bounded, $|\omega_i|$ is bounded uniformly in i. We can show that $\mu_{u_i}\exp(\mu_{u_i} + \omega_i)/(n\alpha\sigma_u^2) = o_p(1)$ we call this result R1. If R1 fails for some $\eta > 0$, then it can be shown that "for some

$i \in \{1, 2, ..., m\}$, both $\mu_{u_i} > 0$ and $\mu_{u_i} \exp(\mu_{u_i} + \omega_i)/(n\alpha\sigma_u^2) > \eta$, for all σ_u^2, β_1^r, and β_1 satisfying $\sigma_u^2, |\beta_0^r|, |\beta_k| \leq C_1, k = 2, ..., p$, and $\sigma_u^2 > C_1^{-1}$." Since $|\omega_i|$ is bounded then $\mu_{u_i} > \log n + O_p(\log\log n)$, and $\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})$ is bounded away from zero and infinity uniformly in $\boldsymbol{\beta}_{-1}^0 \in [-C_1, C_1]^{p-1}$, so for this i, the term μ_{u_i} in the right-hand side of (S4.204) is of order n^c for some $c > 0$. Hence for this i, $\exp(\beta_1^{r0} + U_i) = O_p(n^c)$, and so the probability that the latter bound holds for some $i \in \{1, 2, ..., m\}$ must itself be bounded away from zero along an infinite subsequence of values of n diverging to infinity; call this result R2. Since $P\left[\max_{1 \leq i \leq m} \exp(U_i) > \exp\{C(\log n)^{1/2}\}\right] \to 0$. Since $\exp\{C(\log n)^{1/2}\}$ is of strictly less order that n^c for any $c > 0$, property R2 is violated, and so R1 must be true.

Combining R1 and (S4.202) we have

$$\exp(\beta_1^{r0} - \beta_1^r + U_i - \underline{\widehat{\mu}}_i + \underline{\widehat{\lambda}}_{u_i}/2)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) = 1 + o_p(1) \quad \text{(S4.205)}$$

uniformly in $1 \leq i \leq m$. Taking logarithms of both sides of (S4.205), we have

$$\beta_1^{r0} - \beta_1^r + U_i - \underline{\widehat{\mu}}_i + \underline{\widehat{\lambda}}_{u_i}/2 + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\} = o_p(1) \quad \text{(S4.206)}$$

uniformly in $1 \leq i \leq m$. By (S4.203), we have

$$\begin{aligned}\underline{\widehat{\lambda}}_{u_i} \;&=\; [\sigma_u^{-2} + n\alpha\{\exp(\beta_1^{r0} - \beta_1^r + U_i - \underline{\widehat{\mu}}_i + \underline{\widehat{\lambda}}_{u_i}/2)\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) \\ &\quad + O_p(n^{1-\rho})\}]^{-1}. \qquad \text{(S4.207)}\end{aligned}$$

uniformly in $1 \le i \le m$, for any $\rho \in (0, 1/2)$. Substituting (S4.205) into (S4.207), we have

$$\underline{\widehat{\lambda}}_{u_i} = [\sigma_u^{-2} + n\alpha\{1 + o_p(1) + O_p(n^{1-\rho})\}]^{-1} = O_p(n^{-1}). \quad \text{(S4.208)}$$

uniformly in $1 \le i \le m$. Combining (S4.206) and (S4.208) we have

$$\underline{\widehat{\mu}}_{u_i} = \beta_1^{r0} - \beta_1^r + U_i + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\} + o_p(1). \quad \text{(S4.209)}$$

uniformly in $1 \le i \le m$. Similarly, we can derive that

$$\underline{\widehat{\lambda}}_{v_j} = [\sigma_v^{-2} + m\alpha\{1 + o_p(1)\} + O_p(m^{1-\rho})]^{-1} = O_p(m^{-1}), \quad \text{(S4.210)}$$

$$\underline{\widehat{\mu}}_{v_j} = \beta_1^{c0} - \beta_1^c + V_j + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\} + o_p(1). \quad \text{(S4.211)}$$

uniformly in $1 \le j \le n$. Combining (S4.190), (S4.192), (S4.209), and (S4.211) we have

$$\beta_1^{r0} - \underline{\widehat{\beta_1^r}} + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\} = o_p(1), \quad \text{(S4.212)}$$

$$\beta_1^{c0} - \underline{\widehat{\beta_1^c}} + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\} = o_p(1). \quad \text{(S4.213)}$$

Combining (S4.176), (S4.197), and (S4.206) we have

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - \mu_{u_i} + \lambda_{u_i}/2) - 1\}\mathbf{X}_{ij,-1} \\
&+\frac{1}{mn}\sum_{j=1}^n\sum_{i=1}^m \alpha\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c - \mu_{v_j} + \lambda_{v_j}/2) - 1\}\mathbf{X}_{ij,-1} \\
= \ &\frac{2\alpha\{1 + o_p(1)\}}{\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})}\{\phi_1(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) + m^{-\rho} + n^{-\rho}\} - \frac{2\alpha}{mn}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1} \\
= \ &\frac{2\alpha\phi_1(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})} - 2\alpha\boldsymbol{\mu}_{X_{-1}} + o_p(1). \quad \text{(S4.214)}
\end{aligned}$$

By Taylor's expansion of $\phi_1(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})/\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})$, we have

$$\frac{\phi_1(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})} = \boldsymbol{\mu}_{X_{-1}} + \Sigma_{X_{-1}}(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1}) + o_p(1). \quad \text{(S4.215)}$$

By the assumption (A9) and (S4.214), (S4.215) implies that the consistency of $\underline{\widehat{\boldsymbol{\beta}}}_{-1}$. Combining (S4.212), (S4.213) with (S4.214), respectively, we obtain the consistency of $\underline{\widehat{\beta^r_1}}$ and $\underline{\widehat{\beta^c_1}}$.

Note that results (S4.183), (S4.208), and (S4.209) imply

$$\underline{\widehat{\sigma}}^2_u = \frac{1}{m}\sum_{i=1}^{m}\left[U_i + \beta_1^{r0} - \beta_1^r + \log\{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})\}\right]^2 + o_p(1),$$

by consistency of $\underline{\widehat{\beta^r_1}}$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1}$, and $m^{-1}\sum_{i=1}^m U_i^2 = (\sigma_u^2)^0 + O_p(m^{-1/2})$, further we have

$$\underline{\widehat{\sigma}}^2_u = (\sigma_u^2)^0 + o_p(1), \quad \text{(S4.216)}$$

which implies $\underline{\widehat{\sigma}}^2_u$ is consistent.

For the consistency of $\underline{\widehat{\sigma}}^2_v$, it can be derived analogously:

$$\underline{\widehat{\sigma}}^2_v = \frac{1}{n}\sum_{j=1}^{n} V_j^2 + o_p(1) = (\sigma_v^2)^0 + o_p(1). \quad \text{(S4.217)}$$

Since $\beta_1^{r0} = \beta_1^0 + (\sigma_v^2)^0/2$ and $\beta_1^{c0} = \beta_1^0 + (\sigma_u^2)^0/2$, we have

$$\beta_1^0 = \frac{1}{2}\left\{\beta_1^{r0} - \frac{(\sigma_v^2)^0}{2} + \beta_1^{c0} - \frac{(\sigma_u^2)^0}{2}\right\} \quad \text{(S4.218)}$$

and can define the estimator of β_1^0 by

$$\underline{\widehat{\beta}}_1 = \frac{1}{2}\left(\underline{\widehat{\beta^r_1}} - \frac{\underline{\widehat{\sigma}}^2_v}{2} + \underline{\widehat{\beta^c_1}} - \frac{\underline{\widehat{\sigma}}^2_u}{2}\right). \quad \text{(S4.219)}$$

Equations (S4.212), (S4.213), (S4.216)-(S4.219) imply that

$$\widehat{\underline{\beta}}_{-1} \quad = \quad \beta_1^0 + o_p(1),$$

and β_1^0 can be estimated consistently.

S4.2 Convergence rate for variational estimators

Since (S4.175) and the results of consistency, we can define $\Delta_{i,1}$ by

$$\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r) \\ = \quad & \phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\exp(\beta_1^{r0} - \beta_1^r + U_i + \Delta_{i,1}), \end{aligned} \tag{S4.220}$$

uniformly in $1 \le i \le m$; similarly, we also can define $\Delta'_{j,1}$ by

$$\frac{1}{m}\sum_{i=1}^{m} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c) = \phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_1)\exp(\beta_1^{c0} - \beta_1^c + V_j + \Delta'_{j,1}),$$

uniformly in $1 \le j \le m$. By Taylor's expansion of $\exp(\Delta_{i,1})$ and (S4.220), we have

$$\begin{aligned} & \frac{1}{n}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r - U_i) - \phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\exp(\beta_1^{r0} - \beta_1^r) \\ & = \phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\exp(\beta_1^{r0} - \beta_1^r)\{\Delta_{i,1} + \frac{1}{2}\Delta_{i,1}^2 + O_p(\Delta_{i,1}^3)\}. \end{aligned}$$

uniformly in $1 \le i \le m$. Therefore,

$$\begin{aligned} & \Delta_{i,1} + \frac{1}{2}\Delta_{i,1}^2 + O_p(\Delta_{i,1}^3) \\ = \quad & \frac{1}{n\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})}\sum_{j=1}^{n} Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1 \end{aligned} \tag{S4.221}$$

uniformly in $1 \le i \le m$. By (S4.221) and the following fact, for each $\epsilon > 0$

$$E\left\{\frac{1}{n\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})}\sum_{j=1}^{n} Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1\right\}^2 = O_p(n^{\epsilon-1})$$

uniformly in $1 \le i \le m$, we deduce that $\Delta_{i,1}$ satisfies, for all $C > 0$ and each $\rho \in (0, 1/2)$,

$$\max_{1\le i\le m}\sup_{\|\beta_1^0\|,\|\beta_1\|\le C_1}|\Delta_{i,1}| = O_p(n^{-\rho}), \tag{S4.222}$$

Similarly, we can deduce that $\Delta'_{j,1}$ satisfies, for all $C > 0$ and each $\rho \in (0, 1/2)$,

$$\max_{1\le j\le n}\sup_{\|\beta_1^0\|,\|\beta_1\|\le C_1}|\Delta'_{j,1}| = O_p(m^{-\rho}). \tag{S4.223}$$

Squaring both sides of (S4.221),

$$\begin{aligned}\Delta_{i,1}^2 \;&=\; \left\{\frac{1}{n\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})}\sum_{j=1}^{n} Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1\right\}^2 \\ &\quad +O_p(n^{\epsilon-3/2})\end{aligned}$$

uniformly in $1 \leq i \leq m$, for each $\epsilon > 0$.

$$
\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m} E(\Delta_{i,1}^2) \\
= \;& \frac{1}{m}\sum_{i=1}^{m} E\left\{\frac{1}{n\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}\sum_{j=1}^{n} Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)-1\right\}^2 \\
= \;& \frac{1}{n^2\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})^2} \\
& \times E\left[\sum_{j=1}^{n}\{Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)-\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})\}\right]^2 \\
= \;& \frac{1}{n\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})^2}[\frac{1}{\alpha}\phi(2\boldsymbol{\beta}_{-1}^0-2\boldsymbol{\beta}_{-1})\exp\{(\sigma_v^2)^0\}-\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})^2] \\
= \;& O_p(n^{-1}),
\end{aligned}
$$

we also have

$$
\begin{aligned}
\frac{1}{m}\sum_{i=1}^{m}\Delta_{i,1} \;=\;& \frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i) \\
& -1+O_p(n^{-1}).
\end{aligned}
$$

Since

$$
\begin{aligned}
& E\left\{\frac{1}{mn\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)-1\right\}^2 \\
= \; & \frac{1}{m^2n^2\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})^2}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}^2\exp(-2\beta_1^{r0}-2\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-2U_i) \\
& +\frac{1}{m^2n^2\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})^2} \\
& \times\sum_{i=1}^{m}\sum_{k\neq i}^{m}\sum_{j=1}^{n}Y_{ij}Y_{kj}\exp(-2\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-\mathbf{X}_{kj,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i-U_k) \\
& +\frac{1}{m^2n^2\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})^2} \\
& \times\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l\neq j}^{n}Y_{ij}Y_{il}\exp(-2\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-\mathbf{X}_{il,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-2U_i) \\
& +\frac{1}{m^2n^2\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})^2} \\
& \times\sum_{i=1}^{m}\sum_{k\neq i}^{n}\sum_{j=1}^{n}\sum_{l\neq j}^{n}Y_{ij}Y_{kl}\exp(-2\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-\mathbf{X}_{kl,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i-U_k)-1 \\
= \; & \frac{\phi(2\boldsymbol{\beta}^0_{-1}-2\boldsymbol{\beta}_{-1})}{mn\phi(\boldsymbol{\beta}^0_{-1}-\boldsymbol{\beta}_{-1})^2}\frac{1}{\alpha}\exp\{(\sigma_v^2)^0\}+\frac{m-1}{mn}\exp\{(\sigma_v^2)^0\} \\
& +\frac{n-1}{mn}+\frac{(n-1)(m-1)}{mn}-1 \\
= \; & \frac{1}{n}[\exp\{(\sigma_v^2)^0\}-1]+O_p(m^{-1}n^{-1}),
\end{aligned}
$$

we can deduce that

$$
\frac{1}{m}\sum_{i=1}^{m}\Delta_{i,1}=O_p\{(mn)^{-1/2}+n^{-1/2}\}. \tag{S4.224}
$$

The following property can be derived analogously:

$$
\frac{1}{n}\sum_{j=1}^{n}\Delta'_{j,1}=O_p\{(mn)^{-1/2}+m^{-1/2}\}. \tag{S4.225}
$$

Furthermore, this fact

$$
\begin{aligned}
E\left(\frac{1}{m}\sum_{i=1}^{m} U_i\Delta_{i,1}\right)^2 &= \frac{\phi(2\boldsymbol{\beta}_{-1}^0 - 2\boldsymbol{\beta}_{-1})}{mn\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})^2}\frac{1}{\alpha}(\sigma_u^2)^0\exp\{(\sigma_v^2)^0\} - \frac{1}{mn}(\sigma_u^2)^0 \\
&= O_p(m^{-1}n^{-1}), \qquad \text{(S4.226)}
\end{aligned}
$$

implies that

$$
\frac{1}{m}\sum_{i=1}^{m} U_i\Delta_{i,1} = O_p(m^{-1/2}n^{-1/2}).
$$

The same property is true for $\sum_{j=1}^{n} V_j\Delta'_{j,1}/n$,

$$
\frac{1}{n}\sum_{j=1}^{n} V_j\Delta'_{j,1} = O_p(m^{-1/2}n^{-1/2}).
$$

Based on the notations of (S4.220) and (S4.221), (S4.198) is equivalent to

$$
\begin{aligned}
&\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\exp(\beta_1^{r0} - \beta_1^r + U_i + \Delta_{i,1} - \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2) \\
= \;& 1 + \frac{\underline{\widehat{\mu}}_{u_i}}{n\alpha\sigma_u^2}, \qquad \text{(S4.227)}
\end{aligned}
$$

and (S4.200) is equivalent to

$$
\begin{aligned}
&\phi(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1})\exp(\beta_1^{c0} - \beta_1^c + V_j + \Delta'_{j,1} - \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2) \\
= \;& 1 + \frac{\underline{\widehat{\mu}}_{v_j}}{m\alpha\sigma_v^2}. \qquad \text{(S4.228)}
\end{aligned}
$$

Taking logarithms of both sides of (S4.227) and (S4.228), respectively, we have

$$
\begin{aligned}
& \log\{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})\} + \beta_1^{r0} - \beta_1^r + U_i + \Delta_{i,1} - \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2 \\
= \;& \frac{\underline{\widehat{\mu}}_{u_i}}{n\alpha\sigma_u^2} + o_p(n^{-1}), \qquad \text{(S4.229)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \log\{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})\} + \beta_1^{c0} - \beta_1^c + V_j + \Delta'_{j,1} - \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2 \\
= \;& \frac{\underline{\widehat{\mu}}_{v_j}}{m\alpha\sigma_v^2} + o_p(m^{-1}), \qquad \text{(S4.230)}
\end{aligned}
$$

Adding (S4.229) over i and adding (S4.230) over j, β_0^r, β_0^c, and β_1 satisfy the following equations,

$$
\begin{aligned}
& \beta_1^{r0} - \beta_1^r + \log\{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})\} \\
= \;& -\frac{1}{m}\sum_{i=1}^{m}(U_i + \Delta_{i,1} - \underline{\widehat{\mu}}_{u_i} + \underline{\widehat{\lambda}}_{u_i}/2), \qquad \text{(S4.231)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \beta_1^{c0} - \beta_1^c + \log\{\phi(\boldsymbol{\beta}^0_{-1} - \boldsymbol{\beta}_{-1})\} \\
= \;& -\frac{1}{n}\sum_{j=1}^{n}(V_j + \Delta'_{j,1} - \underline{\widehat{\mu}}_{v_j} + \underline{\widehat{\lambda}}_{v_j}/2). \qquad \text{(S4.232)}
\end{aligned}
$$

By $m^{-1}\sum_{i=1}^m U_i = O_p(m^{-1/2})$, (S4.190), (S4.224), (S4.225), and (S4.208), we have

$$
\begin{aligned}
\beta_1^{r0} - \widehat{\underline{\beta}_1^r} + \log\{\phi(\boldsymbol{\beta}^0_{-1} - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\} &= O_p(m^{-1/2} + n^{-1/2}), \qquad \text{(S4.233)} \\
\beta_1^{c0} - \widehat{\underline{\beta}_1^c} + \log\{\phi(\boldsymbol{\beta}^0_{-1} - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\} &= O_p(m^{-1/2} + n^{-1/2}). \qquad \text{(S4.234)}
\end{aligned}
$$

Setting equation (S4.176) equal to zero we have

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}-\mu_{u_i}+\lambda_{u_i}/2)-1\} \tag{S4.235}$$
$$+\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c}-\mu_{v_j}+\lambda_{v_j}/2)-1\}=0.$$

Combining (S4.229), (S4.230), (S4.234), and (S4.235) we have

$$\begin{aligned}
0 &= \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r}-\mu_{u_i}+\lambda_{u_i}/2)-1\}\\
&\quad+\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c}-\mu_{v_j}+\lambda_{v_j}/2)-1\}\\
&= \frac{\exp(O_p\{(mn)^{\epsilon-1/2}+n^{\epsilon-1}+n^{-1}\})}{\phi(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}\\
&\quad\times\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)\\
&\quad+\frac{\exp(O_p\{(mn)^{\epsilon-1/2}+m^{\epsilon-1}+m^{-1}\})}{\phi(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}\\
&\quad\times\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}Y_{ij}\exp(-\beta_1^{c0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-V_j)\\
&\quad-2\boldsymbol{\mu}_{X_{-1}+O_p(m^{-1/2}n^{-1/2})}\\
&= \frac{1+O_p\{(mn)^{\epsilon-1/2}+n^{\epsilon-1}\}}{\phi(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}\{\phi_1(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})+O_p(n^{-1/2})\}\\
&\quad+\frac{1+O_p\{(mn)^{\epsilon-1/2}+m^{\epsilon-1}\}}{\phi(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}\{\phi_1(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})+O_p(m^{-1/2})\}-2\boldsymbol{\mu}_{X_{-1}}\\
&= \frac{2\phi_1(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}{\phi(\boldsymbol{\beta}_{-1}^{0}-\boldsymbol{\beta}_{-1})}-2\boldsymbol{\mu}_{X_{-1}}+O_p(m^{-1/2}+n^{-1/2}),
\end{aligned}$$

which implies that

$$\underline{\widehat{\boldsymbol{\beta}}}_{-1}-\boldsymbol{\beta}_{-1}^{0} \;=\; O_p(m^{-1/2}+n^{-1/2}). \tag{S4.236}$$

Combining (S4.233), (S4.234), and (S4.236) we have

$$\widehat{\underline{\beta}_1^r} - \beta_1^{r0} = O_p(m^{-1/2} + n^{-1/2}), \tag{S4.237}$$

$$\widehat{\underline{\beta}_1^c} - \beta_1^{c0} = O_p(m^{-1/2} + n^{-1/2}). \tag{S4.238}$$

By (S4.208), (S4.229), and (S4.233), we have

$$\widehat{\underline{\mu}}_{u_i} = U_i + \Delta_{i,1} + O_p(m^{-1/2} + n^{-1/2}), \tag{S4.239}$$

uniformly in $1 \leq i \leq m$. By (S4.222) and (S4.239), we deduce that

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^{m} \widehat{\underline{\mu}}_{u_i}^2 &= \frac{1}{m}\sum_{i=1}^{m} U_i^2 + O_p(m^{-1/2} + n^{-1/2}) \\ &= (\sigma_u^2)^0 + O_p(m^{-1/2} + n^{-1/2}). \end{aligned} \tag{S4.240}$$

By (S4.210), (S4.230), and (S4.234), we deduce that

$$\widehat{\underline{\mu}}_{v_j} = V_j + \Delta'_{j,1} + O_p(m^{-1/2} + n^{-1/2}), \tag{S4.241}$$

uniformly in $1 \leq j \leq n$. By (S4.223) and (S4.241), we deduce that

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n} \widehat{\underline{\mu}}_{v_j}^2 &= \frac{1}{n}\sum_{j=1}^{n} V_j^2 + O_p(m^{-1/2} + n^{-1/2}) \\ &= (\sigma_v^2)^0 + O_p(m^{-1/2} + n^{-1/2}). \end{aligned} \tag{S4.242}$$

Combining (S4.183), (S4.208), and (S4.240) we deduce that

$$\widehat{\underline{\sigma}}_u^2 = (\sigma_u^2)^0 + O_p(m^{-1/2} + n^{-1/2}), \tag{S4.243}$$

and combining (S4.184), (S4.210), and (S4.242) we deduce that

$$\widehat{\underline{\sigma}}_v^2 = (\sigma_v^2)^0 + O_p(m^{-1/2} + n^{-1/2}). \tag{S4.244}$$

Together, (S4.218), (S4.219), (S4.237), (S4.238), (S4.243), and (S4.244) give

$$\begin{aligned}\widehat{\underline{\beta}}_1 - \beta_1^0 &= \frac{1}{2}\left\{\widehat{\underline{\beta}_1^r} - \beta_1^{r0} - \frac{\widehat{\underline{\sigma}}_v^2 - (\sigma_v^2)^0}{2} + \widehat{\underline{\beta}_1^c} - \beta_1^{c0} - \frac{\widehat{\underline{\sigma}}_u^2 - (\sigma_u^2)^0}{2}\right\} \\ &= O_p(m^{-1/2} + n^{-1/2}).\end{aligned}$$

This completes the proof of consistency at the rate $m^{-1/2} + n^{-1/2}$ for the Gamma regression model in Theorem 1.

S5 Proof of Theorem 2 for Gamma regression model

For Gamma case, the proof of Theorem 2 includes 3 steps:

Step1. initial approximation to $\widehat{\underline{\beta}_1^r} - \beta_1^{r0}$, $\widehat{\underline{\beta}_1^c} - \beta_1^{c0}$, and $\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0$ in §S5.1;

Step2. Final approximations to $\widehat{\underline{\sigma}}_u^2 - (\sigma_u^2)^0$ and $\widehat{\underline{\sigma}}_v^2 - (\sigma_v^2)^0$ in §S5.2;

Step3. Final approximations to $\widehat{\underline{\beta}}_1 - \beta_1^0$ and $\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0$ in §S5.3.

S5.1 Initial approximation to $\widehat{\underline{\beta}_1^r} - \beta_1^{r0}, \widehat{\underline{\beta}_1^c} - \beta_1^{c0}$ and $\widehat{\underline{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0$

Denote $\boldsymbol{\mu}_{\mathbf{X}_{-1}}$ and $\Sigma^2_{\mathbf{X}_{-1}}$ the mean and variance of $\mathbf{X}_{-1}$. Equating (S4.176) to zero we have

$$\begin{aligned}&\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\underline{\boldsymbol{\beta}}^r} - \widehat{\underline{\mu}}_{u_i} + \widehat{\underline{\lambda}}_{u_i}/2) \\ &+\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\widehat{\underline{\boldsymbol{\beta}}^c} - \widehat{\underline{\mu}}_{v_j} + \widehat{\underline{\lambda}}_{v_j}/2) \\ &-\frac{2}{mn}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1} = 0. \qquad \text{(S5.245)}\end{aligned}$$

By (S4.236), (S5.245), and Taylor expansions of $\exp\{\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\}$

$$\begin{aligned}
& \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}Y_{ij}\exp(-\widehat{\underline{\beta}_1^r}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\mu}}_{u_i}+\underline{\widehat{\lambda}}_{u_i}/2)\\
& +\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}\exp(-\widehat{\underline{\beta}_1^r}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\mu}}_{u_i}+\underline{\widehat{\lambda}}_{u_i}/2)\\
& \times\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\\
& +\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}Y_{ij}\exp(-\widehat{\underline{\beta}_1^c}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\mu}}_{v_j}+\underline{\widehat{\lambda}}_{v_j}/2)\\
& +\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}Y_{ij}\exp(-\widehat{\underline{\beta}_1^c}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\mu}}_{v_j}+\underline{\widehat{\lambda}}_{v_j}/2)\\
& \times\mathbf{X}_{ij,-1}\mathbf{X}_{ij-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\\
= \;& \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}+O_p\{\|\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1}\|^2\}. \qquad \text{(S5.246)}
\end{aligned}$$

Since (S4.229) imply that

$$\begin{aligned}
\underline{\widehat{\mu}}_{u_i}\left(1+\frac{1}{n\alpha\sigma_u^2}\right) &= U_i+\Delta_{i,1}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}+\beta_1^{r0}-\widehat{\underline{\beta}_1^r}\\
&\quad+\log\{\phi(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\}, \qquad \text{(S5.247)}
\end{aligned}$$

uniformly in $1\le i\le m$, for each $\epsilon>0$, adding (S5.247) over i and dividing both sides by m, we have

$$\begin{aligned}
& \frac{1}{m}\sum_{i=1}^{m}\underline{\widehat{\mu}}_{u_i}\left(1+\frac{1}{n\alpha\sigma_u^2}\right) \qquad \text{(S5.248)}\\
= \;& \frac{1}{m}\sum_{i=1}^{m}[U_i+\Delta_{i,1}+\frac{1}{2}\underline{\widehat{\lambda}}_{u_i}+\beta_1^{r0}-\widehat{\underline{\beta}_1^r}+\log\{\phi(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\}].
\end{aligned}$$

Substituting the (S4.190) into (S5.248) we deduce that

$$\widehat{\underline{\beta}_1^r}=\beta_1^{r0}+\bar{U}+\bar{\Delta}_{i,1}+\log\{\phi(\boldsymbol{\beta}_{-1}^{0}-\underline{\widehat{\boldsymbol{\beta}}}_{-1})\}. \qquad \text{(S5.249)}$$

By (S5.247), (S4.208), and (S4.231), we have

$$\underline{\widehat{\mu}}_{u_i} = (U_i + \Delta_{i,1} - \bar{U} - \bar{\Delta}_{i,1})\left\{1 - \frac{1}{n\alpha\sigma_u^2} + \frac{2}{n^2\alpha^2\sigma_u^4}\right\}. \quad \text{(S5.250)}$$

Combining (S4.210), (S4.211), and (S4.232) we can derive analogous properties,

$$\widehat{\underline{\beta}_1^c} = \beta_1^{c0} + \bar{V} + \bar{\Delta}'_{j,1} + \log\{\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\}, \quad \text{(S5.251)}$$

$$\underline{\widehat{\mu}}_{v_j} = (V_j + \Delta'_{j,1} - \bar{V} - \bar{\Delta}'_{j,1})\{1 + O_p(m^{-1})\}. \quad \text{(S5.252)}$$

Substituting the (S5.249), (S5.250), (S5.251), and (S5.252) into (S5.246) we have

$$\begin{aligned}
& \frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i - \Delta_{i,1} - \frac{U_i}{n\alpha\sigma_u^2}) \\
& + \frac{\exp\{O_p(n^{\epsilon-1})\}}{mn\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})}\sum_{i=1}^m\sum_{j=1}^n Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i - \Delta_{i,1}) \\
& \times \mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
& + \frac{1}{mn\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j - \Delta'_{j,1} - \frac{V_j}{m\alpha\sigma_v^2}) \\
& + \frac{\exp\{O_p(m^{\epsilon-1})\}}{mn\phi(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1})}\sum_{i=1}^m\sum_{j=1}^n Y_{ij}\exp(-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j - \Delta'_{j,1}) \\
& \times \mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^0 - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= \; & \frac{2}{mn}\sum_{i=1}^m\sum_{j=1}^n \mathbf{X}_{ij,-1},
\end{aligned}$$

by $\exp\{O_p(n^{\epsilon-1})\} = 1 + O_p(n^{\epsilon-1})$ and $\exp(-\Delta_{i,1}) = 1 - \Delta_{i,1} + O_p(n^{\epsilon-1})$,

$$\begin{aligned}
& \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i)\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
& + \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j)\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
& + O_p\{m^{\epsilon-3/2} + n^{\epsilon-3/2}\} \\
& + \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}^{\mathrm{T}} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i)(1 - \Delta_{i,1}) + O_p(m^{\epsilon-3/2}) \\
& + \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{X}_{ij,-1}^{\mathrm{T}} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j)(1 - \Delta'_{j,1}) + O_p(n^{\epsilon-3/2}) \\
= \; & \phi(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1})\frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{X}_{ij,-1}. \qquad \text{(S5.253)}
\end{aligned}$$

By the law of large numbers, we deduce that

$$\begin{aligned}
& \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i)\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}} \\
= \; & \boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{X}_{-1}} + o_p(1), \qquad \text{(S5.254)}
\end{aligned}$$

and

$$\begin{aligned}
& \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - U_i)\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}} \\
= \; & \boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{X}_{-1}} + o_p(1). \qquad \text{(S5.255)}
\end{aligned}$$

Using Taylor's expansion to $\phi(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1})$ we have

$$\begin{aligned}
& \phi(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
= \; & 1 + \boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1}) \\
& + (\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1})^{\mathrm{T}}(\boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{X}_{-1}})(\boldsymbol{\beta}_{-1}^{0} - \underline{\widehat{\boldsymbol{\beta}}}_{-1}). \qquad \text{(S5.256)}
\end{aligned}$$

Substituting (S5.254), (S5.255), and (S5.256) into (S5.253) we have

$$\begin{aligned} & 2\mathbf{\Sigma}_{\mathbf{X}_{-1}}(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) \\ = \; & T_1 + T_2 + T_3 + T_4 + o_p\{m^{-1/2}n^{-1/2} + \|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\}. \end{aligned} \tag{S5.257}$$

where

$$\begin{aligned} T_1 &= \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \mathbf{X}_{ij,-1}\} - \frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{m}\sum_{i=1}^m S_1, \\ T_2 &= \frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j) - \mathbf{X}_{ij,-1}\} - \frac{\boldsymbol{\mu}_{X_{-1}}}{n}\sum_{j=1}^n S_2, \\ T_3 &= -\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \boldsymbol{\mu}_{\mathbf{X}_{-1}}\}S_1 + \frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{m}\sum_{i=1}^m S_1^2, \\ T_4 &= -\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j) - \boldsymbol{\mu}_{\mathbf{X}_{-1}}\}S_2 + \frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{n}\sum_{j=1}^n S_2^2. \end{aligned}$$

Notice that

$$E(T_3) = O_p\{n^{-1}\|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}\|^2\}, \tag{S5.258}$$

$$E(T_4) = O_p\{m^{-1}\|\underline{\widehat{\boldsymbol{\beta}}}_{1} - \boldsymbol{\beta}^0_{-1}\|^2\}. \tag{S5.259}$$

Combining (S5.257), (S5.258), and (S5.259) we deduce that

$$2\mathbf{\Sigma}_{\mathbf{X}_{-1}}(\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1}) = T_1 + T_2 + o_p\{m^{-1/2}n^{-1/2} + \|\underline{\widehat{\beta}}_1 - \beta_1^0\|^2\}. \tag{S5.260}$$

Observe that $E(T_1) = E(T_2) = 0$. The variance of T_1 is

$$
\begin{aligned}
& E(T_1 T_1^{\mathrm{T}}) \\
= & \left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \mathbf{X}_{ij,-1}\} - \frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{m}\sum_{i=1}^{m} S_1\right] \\
& \times\left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \mathbf{X}_{ij,-1}\} - \frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{m}\sum_{i=1}^{m} S_1\right]^{\mathrm{T}} \\
= & \left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \mathbf{X}_{ij,-1}\}\right. \\
& \left.-\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_0^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - \boldsymbol{\mu}_{\mathbf{X}_{-1}}]\right] \\
& \times\left[\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - \mathbf{X}_{ij,-1}\}\right. \\
& \left.-\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - \boldsymbol{\mu}_{\mathbf{X}_{-1}}]\right]^{\mathrm{T}} \\
= & T_{11} + T_{12} + T_{13}, \qquad \text{(S5.261)}
\end{aligned}
$$

where

$$
\begin{aligned}
T_{11} = & \frac{1}{mn}E\Big[\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - 1\}^2\mathbf{X}_{ij,-1}\mathbf{X}_{ij,-1}^{\mathrm{T}} \\
& +\{\frac{1}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1\}^2\boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} \\
& -2\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - 1\} \\
& \times\{\frac{1}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1\}\mathbf{X}_{ij,-1}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}}\Big] \\
= & \frac{1}{mn}[\exp\{(\sigma_v^2)^0\} - 1]\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} + \frac{\exp\{(\sigma_v^2)^0\}}{mn\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} \qquad \text{(S5.262)}
\end{aligned}
$$

$$\begin{aligned}
T_{12} &= \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{l\neq j}^{n} E[\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}-U_i)-\mathbf{X}_{ij,-1}\} \\
&\quad -\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)+\boldsymbol{\mu}_{\mathbf{X}_{-1}}\} \\
&\quad \times\{\mathbf{X}_{il,-1}Y_{il}\exp(-\mathbf{X}_{il}^{\mathrm{T}}\boldsymbol{\beta}^{r0}-U_i)-\mathbf{X}_{il,-1} \\
&\quad -\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}Y_{il}\exp(-\beta_1^{r0}-\mathbf{X}_{il,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)+\boldsymbol{\mu}_{\mathbf{X}_{-1}}\}] \\
&= 0, \qquad\qquad (\text{S5.263}) \\
T_{13} &= \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{k\neq i}^{m}\sum_{j=1}^{n} E[\{\mathbf{X}_{ij,-1}Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0}-U_i)-\mathbf{X}_{ij,-1} \\
&\quad -\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0}-\mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1}-U_i)+\boldsymbol{\mu}_{\mathbf{X}_{-1}}\} \\
&\quad \times\{\mathbf{X}_{kj,-1}Y_{kj}\exp(-\mathbf{X}_{kj}\boldsymbol{\beta}^{r0}-U_k)-\mathbf{X}_{kj,-1}\} \\
&\quad -\frac{\boldsymbol{\mu}_{\mathbf{X}_{-1}}}{\phi(\boldsymbol{\beta}_{-1}^0-\boldsymbol{\beta}_{-1})}Y_{kj}\exp(-\beta_1^{r0}-\mathbf{X}_{kj,-1}\boldsymbol{\beta}_{-1}-U_k)+\boldsymbol{\mu}_{\mathbf{X}_{-1}}\}] \\
&= 0. \qquad\qquad (\text{S5.264})
\end{aligned}$$

Combining (S5.261)-(S5.264) we deduce that

$$E(T_1T_1^{\mathrm{T}}) = \frac{\exp\{(\sigma_v^2)^0\}-1}{mn}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}+\frac{\exp\{(\sigma_v^2)^0\}}{mn\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}. \quad (\text{S5.265})$$

The variance of T_2 can be calculated analogously,

$$E(T_2T_2^{\mathrm{T}}) = \frac{\exp\{(\sigma_u^2)^0\}-1}{mn}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}+\frac{\exp\{(\sigma_u^2)^0\}}{mn\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}. \quad (\text{S5.266})$$

Furthermore, we have

$$T_1 = O_p\{(mn)^{-1/2}\} \text{ and } T_2 = O_p\{(mn)^{-1/2}\}.$$

By (S4.231), (S4.232), and (S5.256),

$$\begin{aligned}\beta_1^r - \beta_1^{r0} &= \phi_1^{\mathrm{T}}(\mathbf{0})(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) + O_p\{(mn)^{-1/2}\} \\ &\quad + \frac{1}{m}\sum_{i=1}^m \{U_i + \Delta_{i,1} - \widehat{\underline{\mu}}_{u_i} + \widehat{\underline{\lambda}}_{u_i}/2\}, \qquad \text{(S5.267)} \\ \beta_1^c - \beta_1^{c0} &= \phi_1^{\mathrm{T}}(\mathbf{0})(\boldsymbol{\beta}_{-1}^0 - \boldsymbol{\beta}_{-1}) + O_p\{(mn)^{-1/2}\} \\ &\quad + \frac{1}{n}\sum_{j=1}^n \{V_j + \Delta'_{j,1} - \widehat{\underline{\mu}}_{v_j} + \widehat{\underline{\lambda}}_{v_j}/2\}. \qquad \text{(S5.268)}\end{aligned}$$

S5.2 Final approximations to $\sigma_u^2 - (\sigma_u^2)^0$ and $\sigma_v^2 - (\sigma_v^2)^0$

Therefore, squaring both sides of (S4.237) and adding i from 1 to m we have

$$\frac{1}{m}\sum_{i=1}^m \widehat{\underline{\mu}}_{u_i}^2 = \frac{1}{m}\sum_{i=1}^m (U_i + \Delta_{i,1} - \bar{U} - \bar{\Delta}_{i,1})^2 + O_p(n^{-1}).$$

Combining (S4.183), (S4.208), (S4.227), and (S4.237) we have

$$\begin{aligned}\widehat{\underline{\sigma}}_u^2 &= \frac{1}{m}\sum_{i=1}^m (U_i + \Delta_{i,1} - \bar{U} - \bar{\Delta}_{i,1})^2 + O_p\{n^{-1} + (mn)^{-1/2}\} \\ &= \frac{1}{m}\sum_{i=1}^m U_i^2 + \frac{1}{m}\sum_{i=1}^m \Delta_{i,1}^2 + \frac{2}{m}\sum_{i=1}^m U_i\Delta_{i,1} + O_p(m^{-1} + n^{-1}) \\ &= \frac{1}{m}\sum_{i=1}^m U_i^2 + O_p\{(m^{-1/2} + n^{-1/2})^2\}.\end{aligned}$$

Therefore,

$$\widehat{\underline{\sigma}}_u^2 - (\sigma_u^2)^0 = \frac{1}{m}\sum_{i=1}^m \{U_i^2 - (\sigma_u^2)^0\} + O_p\{(m^{-1/2} + n^{-1/2})^2\},$$

which implies the asymptotic normality result of $\underline{\widehat{\sigma}}_u^2$ in Theorem 2. The expression of $\underline{\widehat{\sigma}}_v^2$ can be derived analogously:

$$\underline{\widehat{\sigma}}_v^2 = \frac{1}{n}\sum_{j=1}^{n}(V_j - \Delta'_{j,1} - \bar{V} - \bar{\Delta}'_{j,1})^2 + O_p\{(m^{-1/2} + n^{-1/2})^2\},$$

and

$$\underline{\widehat{\sigma}}_v^2 - (\sigma_v^2)^0 = \frac{1}{n}\sum_{j=1}^{n}\{V_j^2 - (\sigma_v^2)^0\} + O_p\{(m^{-1/2} + n^{-1/2})^2\},$$

which implies the asymptotic normality result of $\underline{\widehat{\sigma}}_v^2$ in Theorem 2.

S5.3 Final approximations to $\underline{\widehat{\beta}}_1 - \beta_1^0$ and $\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0$

Based on (S5.260), we have

$$\begin{aligned} \underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0 &= \frac{1}{2}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}^{-1}T_1 + \frac{1}{2}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}^{-1}T_2 \\ &\quad + o_p\{(mn)^{-1/2} + \|\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}_{-1}^0\|^2\}. \end{aligned} \qquad \text{(S5.269)}$$

Using (S5.265), (S5.266), and the covariance of T_1 and T_2,

$$
\begin{aligned}
& \operatorname{cov}(T_1, T_2) \\
= \; & E\Bigg(\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\Bigg[\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{r0} - U_i) - 1\}\mathbf{X}_{ij,-1} \\
& -\{\frac{1}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{r0} - \mathbf{X}_{ij,-1}^{\mathrm{T}}\boldsymbol{\beta}_{-1} - U_i) - 1\}\boldsymbol{\mu}_{\mathbf{X}_{-1}}\Bigg] \\
& \times\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\Bigg[\{Y_{ij}\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^{c0} - V_j) - 1\}\mathbf{X}_{ij,-1} \\
& -\{\frac{1}{\phi(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})}Y_{ij}\exp(-\beta_1^{c0} - \mathbf{X}_{ij,-1}\boldsymbol{\beta}_{-1} - V_j) - 1\}\boldsymbol{\mu}_{\mathbf{X}_{-1}}\Bigg]^{\mathrm{T}}\Bigg) \\
= \; & \frac{1}{mn}\frac{1}{\alpha}[\boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} + \boldsymbol{\Sigma}_{\mathbf{X}_{-1}} - 2\{\boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}} + o_p(1)\} \\
& +\boldsymbol{\mu}_{\mathbf{X}_{-1}}\boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}}\{1 + \boldsymbol{\mu}_{\mathbf{X}_{-1}}^{\mathrm{T}}(\boldsymbol{\beta}_{-1}^{0} - \boldsymbol{\beta}_{-1})\}] \\
= \; & \frac{1}{mn}\frac{1}{\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} + o_p\{(mn)^{-1}\},
\end{aligned}
$$

Since

$$
\begin{aligned}
& Var(T_1 + T_2) \\
= \; & Var(T_1) + Var(T_2) + 2\operatorname{cov}(T_1, T_2) \\
= \; & \frac{1+\alpha}{mn\alpha}\exp\{(\sigma_v^2)^0\}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} - \frac{1}{mn}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} \\
& +\frac{1+\alpha}{mn\alpha}\exp\{(\sigma_u^2)^0\}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} - \frac{1}{mn}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} + 2\frac{1}{mn\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} \\
= \; & \frac{\exp\{(\sigma_u^2)^0\} + \exp\{(\sigma_v^2)^0\} - 2}{mn}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} \\
& +\frac{\exp\{(\sigma_u^2)^0\} + \exp\{(\sigma_v^2)^0\} + 2}{mn\alpha}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}} \qquad \text{(S5.270)}
\end{aligned}
$$

we can define a covariance matrix $\tilde{\boldsymbol{\Sigma}}$ by

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}} &= \frac{1}{4}\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}^{-1} Var(T_1+T_2)\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}^{-1} \\
&= \left(\frac{1}{4mn}[\exp\{(\sigma_u^2)^0\}+\exp\{(\sigma_v^2)^0\}-2]\right. \\
&\quad \left.+\frac{1}{4mn\alpha}[\exp\{(\sigma_u^2)^0\}+\exp\{(\sigma_v^2)^0\}+2]\right)\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}^{-1}. \qquad \text{(S5.271)}
\end{aligned}$$

Combining (S5.269), (S5.271), and the multivariate central limit theorem we have

$$\sqrt{mn}(\underline{\widehat{\boldsymbol{\beta}}}_{-1}-\boldsymbol{\beta}_{-1}^0) \to N(0,\widetilde{\boldsymbol{\Sigma}})$$

in distribution, as $m,\ n\to\infty$. By (S5.267), (S5.268), and (S5.269), we have

$$\begin{aligned}
& \underline{\widehat{\beta}}_1-\beta_1^0 \qquad \text{(S5.272)} \\
=\ & -(\underline{\widehat{\boldsymbol{\beta}}}_{-1}-\boldsymbol{\beta}_{-1}^0)\mu_X+\frac{1}{2m}\sum_{i=1}^m(U_i+\Delta_{i,1})+\frac{1}{2n}\sum_{j=1}^n(V_i+\Delta_{j,1}') \\
& -\frac{1}{4}\{\underline{\widehat{\sigma}}_u^2-(\sigma_u^2)^0+\underline{\widehat{\sigma}}_v^2-(\sigma_v^2)^0\}+O_p\{(m^{-1/2}+n^{-1/2})^2\} \\
=\ & \frac{1}{2}\bar{V}+\frac{1}{2}\bar{\Delta}_{i,1}-\frac{1}{4n}\sum_{j=1}^n\{V_j^2-(\sigma_v^2)^0\} \\
& +\frac{1}{2}\bar{U}+\frac{1}{2}\bar{\Delta}_{j,1}'-\frac{1}{4m}\sum_{i=1}^m\{U_i^2-(\sigma_u^2)^0\}+O_p\{(m^{-1/2}+n^{-1/2})^2\} \\
:=\ & SS_1+SS_2+SS_3+SS_4+SS_5+SS_6+O_p\{(m^{-1/2}+n^{-1/2})^2\},
\end{aligned}$$

where

$$
\begin{aligned}
SS_1 &= \frac{1}{2}\bar{V}, \\
SS_2 &= \frac{1}{2}\bar{\Delta}_{i,1}, \\
SS_3 &= -\frac{1}{4n}\sum_{j=1}^{n}\{V_j^2 - (\sigma_v^2)^0\}, \\
SS_4 &= \frac{1}{2}\bar{U}, \\
SS_5 &= \frac{1}{2}\bar{\Delta}'_{j,1}, \\
SS_6 &= -\frac{1}{4m}\sum_{i=1}^{m}\{U_i^2 - (\sigma_u^2)^0\}.
\end{aligned}
$$

By the requirements of the multivariate central limit theorem, we need to check the boundedness of covariances of $(SS_1, SS_2, SS_3, SS_4, SS_5, SS_6)$. The diagonal elements can be calculated as follows,

$$
\begin{aligned}
\mathrm{var}(SS_1) &= \frac{1}{4n}(\sigma_v^2)^0, && \text{(S5.273)} \\
\mathrm{var}(SS_2) &= \frac{1}{4n}[\exp\{(\sigma_v^2)^0\} - 1], && \text{(S5.274)} \\
\mathrm{var}(SS_3) &= \frac{1}{8n}\{(\sigma_v^2)^0\}^2, && \text{(S5.275)} \\
\mathrm{var}(SS_4) &= \frac{1}{4m}(\sigma_u^2)^0, && \text{(S5.276)} \\
\mathrm{var}(SS_5) &= \frac{1}{4m}[\exp\{(\sigma_u^2)^0\} - 1], && \text{(S5.277)} \\
\mathrm{var}(SS_6) &= \frac{1}{8m}\{(\sigma_u^2)^0\}^2. && \text{(S5.278)}
\end{aligned}
$$

The covariances can be computed as follows,

$$\begin{aligned}
\mathrm{cov}(SS_1, SS_2) &= \frac{1}{4n}(\sigma_v^2)^0, && \text{(S5.279)}\\
\mathrm{cov}(SS_1, SS_3) &= 0, && \text{(S5.280)}\\
\mathrm{cov}(SS_2, SS_3) &= -\frac{1}{8m}(\sigma_u^4)^0, && \text{(S5.281)}\\
\mathrm{cov}(SS_4, SS_5) &= \frac{1}{4m}(\sigma_u^2)^0, && \text{(S5.282)}\\
\mathrm{cov}(SS_4, SS_6) &= 0, && \text{(S5.283)}\\
\mathrm{cov}(SS_5, SS_6) &= -\frac{1}{8n}\{(\sigma_u^2)^0\}^2 && \text{(S5.284)}
\end{aligned}$$

For simplicity, combining (S5.273)-(S5.284) we define two matrices by

$$\mathbf{\Gamma}_4 = \frac{1}{8}\begin{pmatrix} 2(\sigma_u^2)^0 & 2(\sigma_u^2)^0 & 0 \\ 2(\sigma_u^2)^0 & 2[\exp\{(\sigma_u^2)^0\} - 1] & -\{(\sigma_u^2)^0\}^2 \\ 0 & -\{(\sigma_u^2)^0\}^2 & \{(\sigma_u^2)^0\}^2 \end{pmatrix}. \quad \text{(S5.285)}$$

and

$$\mathbf{\Gamma}_5 = \frac{1}{8}\begin{pmatrix} 2(\sigma_v^2)^0 & 2(\sigma_v^2)^0 & 0 \\ 2(\sigma_v^2)^0 & 2[\exp\{(\sigma_v^2)^0\} - 1] & -\{(\sigma_v^2)^0\}^2 \\ 0 & -\{(\sigma_v^2)^0\}^2 & \{(\sigma_v^2)^0\}^2 \end{pmatrix}, \quad \text{(S5.286)}$$

Combining (S5.272), (S5.285), (S5.286), and the multivariate central limit theorem we deduce that

$$\begin{aligned} & (\sqrt{m}(SS_4, SS_5, SS_6), \sqrt{n}(SS_1, SS_2, SS_3))^{\mathrm{T}} \\ \rightarrow \quad & N\left(\begin{pmatrix} \mathbf{0}_3 \\ \mathbf{0}_3 \end{pmatrix}, \begin{pmatrix} \mathbf{\Gamma}_4 & 0 \\ 0 & \mathbf{\Gamma}_5 \end{pmatrix}\right), \end{aligned} \tag{S5.287}$$

as $m, n \to \infty$. By results (S5.272), (S5.287), we obtain

$$\underline{\widehat{\beta}}_1 - \beta_1^0 \quad = \quad m^{-1/2} 1_3^{\mathrm{T}} Z_7 + n^{-1/2} 1_3^{\mathrm{T}} Z_8 + o_p(m^{-1/2} + n^{-1/2}),$$

where Z_7 and Z_8 are normal distributions with mean zeros and variances $\mathbf{\Gamma}_4$ and $\mathbf{\Gamma}_5$, respectively.

S6 Additional simulation results

S6.1 Simulations for the Gamma regression model with $\delta = 1$

We conduct the simulation studies to assess the performance of the Gamma regression models with crossed random effects as specified in equation (S1.2). Except the case in the mainuscript, we also consider another case, independently generate $\mathbf{X}_{ij,-1}$ from $N(0, I_{8\times 8})$ for $i = 1, ..., m$ and $j = 1, ..., n$, where $I_{8\times 8}$ is an identity matrix. We set $\boldsymbol{\beta}^0 = (-2, 1, 0.5, 1, -1, 0.5, 1, -1, 0.5, 0.8, -0.4, 0.1)$, $(\sigma_u)^0 = 0.5$, and $(\sigma_v)^0 = 0.5$. The additional shape parameter is set as $\alpha = 0.8$. We consider nine different scenarios: $(m, n) \in \{(18, 12), (20, 20), (30, 30), (50, 50), (80, 80), (100, 100), (150, 150), (200, 200), (300, 300)\}$, which imply the nine increasing sample sizes $m*n \in \{216, 400, 900, 2500, 6400, 10000, 22500, 40000, 90000\}$. We report the results based on 500 Monte Carlo studies in Figures S1, and S2.

Theorem 2 reveals that $\underline{\widehat{\boldsymbol{\beta}}}$, $\underline{\widehat{\sigma}}_u$, and $\underline{\widehat{\sigma}}_v$ are each asymptotically normal with means corresponding to the true parameter values. Based on Theorem 2, the approximate 95% confidence intervals for $\beta_1, \beta_j, j = 2, ..., p, \sigma_u^0$, and σ_v^0 are $\underline{\widehat{\beta}}_1 \pm 1.96 * \text{var}(\underline{\widehat{\beta}}_1)$, $\underline{\widehat{\beta}}_j \pm 1.96 * \text{var}(\underline{\widehat{\beta}}_j), j = 2, ..., p$, $\underline{\widehat{\sigma}}_u \pm 1.96 * \text{var}(\underline{\widehat{\sigma}}_u)$, $\underline{\widehat{\sigma}}_v \pm 1.96 * \text{var}(\underline{\widehat{\sigma}}_v)$. These confidence intervals are asymptotically valid, as they rely on studentization using consistent estimators for unknown parameters. We also report the coverage percentages for the 95% confidence intervals in Figure S3.

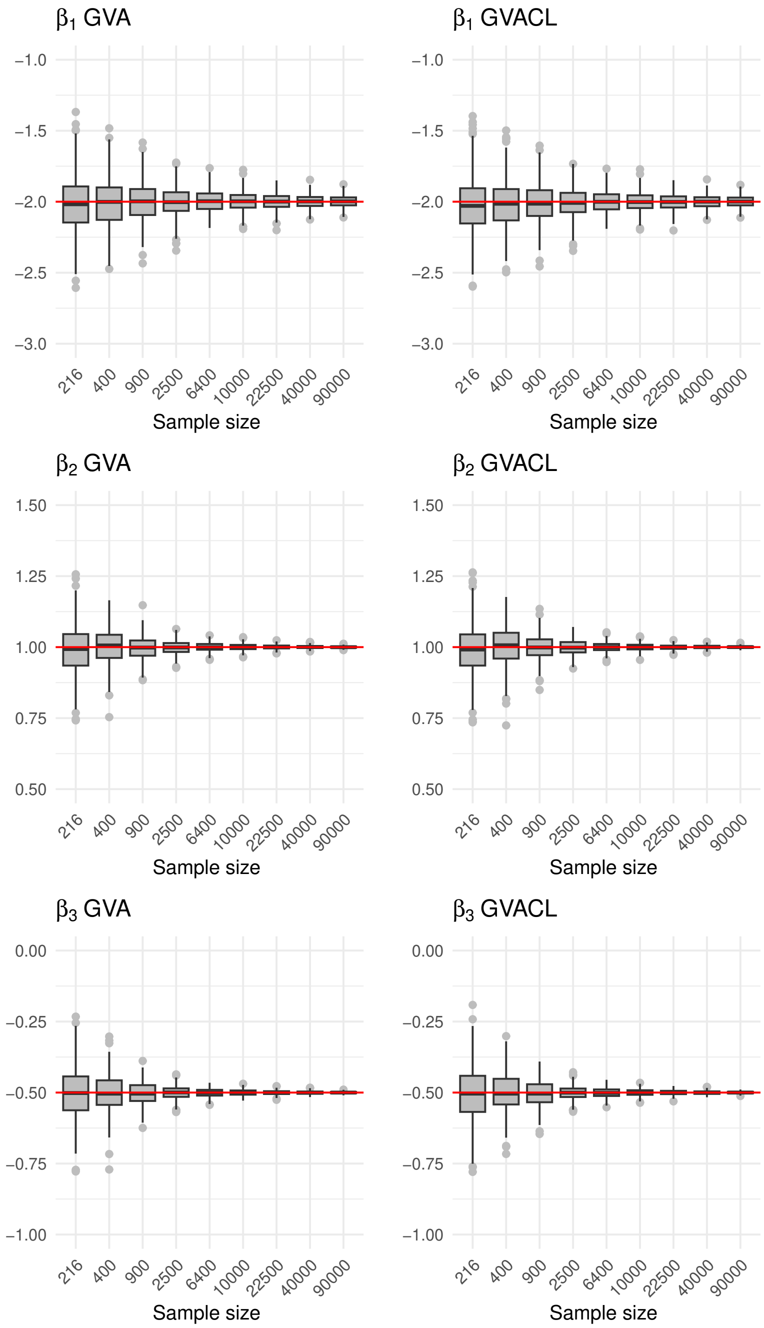

Figure S1: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA and GVACL for the Gamma regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.

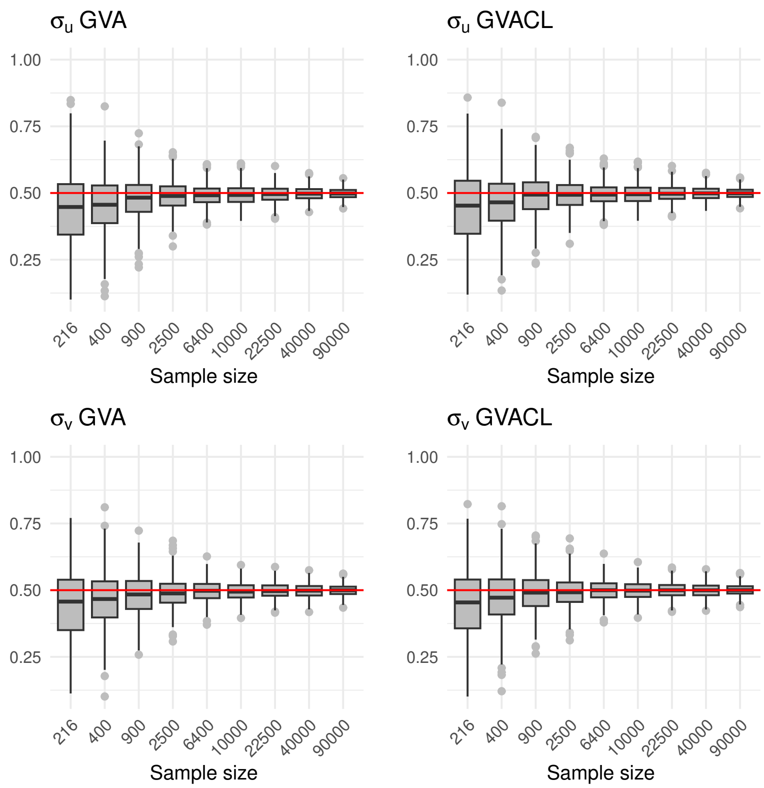

Figure S2: Boxplots of σ_u and σ_v estimates obtained by GVA and GVACL for the Gamma regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.

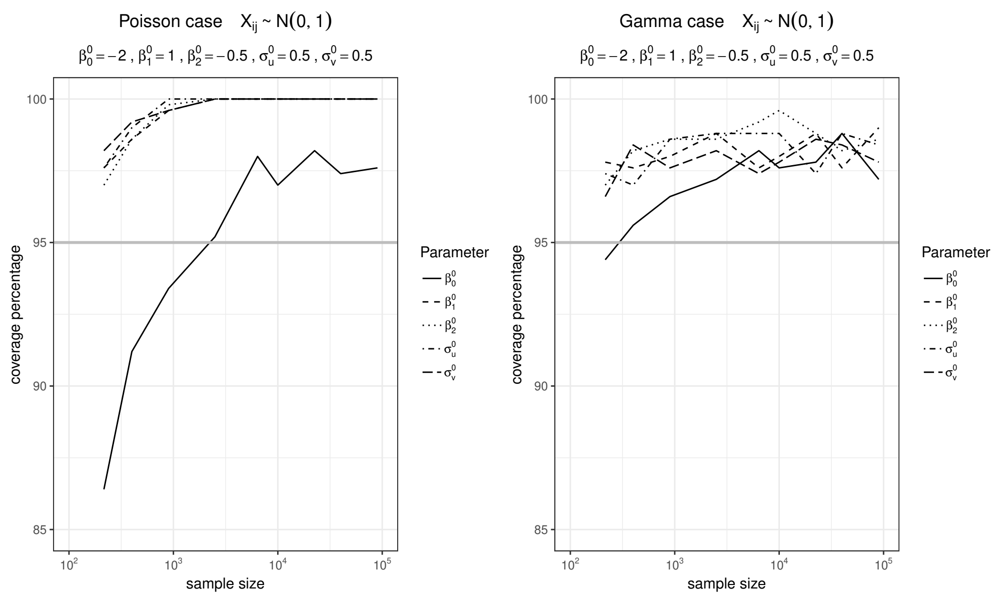

Figure S3: The coverage percentage of the 95% GVACL confidence intervals for the parameters in Poisson and Gamma regression models with crossed random effects based on 500 replications. The 95% percentage is shown as a thick grey horizontal line. The ratio $\log m/\log n$ converges to 1.

S6.2 Simulations for the logistic regression model with $\delta = 1$

We also consider simulation studies for the following logistic regression model with crossed random effects,

$$Y_{ij} = 1 \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Bernoulli}[1/\{1+\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - U_i - V_j)\}],$$

$$U_i \sim N(0, (\sigma_u^2)^0), V_j \sim N(0, (\sigma_v^2)^0), U_i \text{ and } V_j \text{ are independent.}$$

The variational lower bound of the CL for the logistic regression model with crossed random effects is

$$\begin{aligned}
& \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi}) \\
= \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{u_i}) - \int \log(1+\exp(\mathbf{X}_{ij}\boldsymbol{\beta}^r + U_i))\phi(u_i)du_i\} \\
& -\frac{m}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}) + \frac{1}{2}\log\lambda_{u_i} \\
& +\sum_{j=1}^{n}\sum_{i=1}^{m}[Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c + \mu_{v_j}) - \int \log\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c + V_j))\phi(v_j)\}dv_j] \\
& -\frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{j=1}^{n}(\mu_{v_j}^2 + \lambda_{v_j}) + \frac{1}{2}\log\lambda_{v_j} \\
\geq \; & \sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + \mu_{u_i}) - \log(1+\exp(\mathbf{X}_{ij}\boldsymbol{\beta}^r + \mu_i + \lambda_{u_i}/2))\} \\
& -\frac{m}{2}\log(\sigma_u^2) - \frac{1}{2\sigma_u^2}\sum_{i=1}^{m}(\mu_{u_i}^2 + \lambda_{u_i}) + \frac{1}{2}\log\lambda_{u_i} \\
& +\sum_{j=1}^{n}\sum_{i=1}^{m}[Y_{ij}(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c + \mu_{v_j}) - \log\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^c + \mu_{v_j} + \lambda_{v_j}/2))] \\
& -\frac{n}{2}\log(\sigma_v^2) - \frac{1}{2\sigma_v^2}\sum_{j=1}^{n}(\mu_{v_j}^2 + \lambda_{v_j}) + \frac{1}{2}\log\lambda_{v_j} := \underline{\underline{c\ell}}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi}). \quad \text{(S6.288)}
\end{aligned}$$

We consider four scenarios: $(m, n) \in \{(200, 200), (250, 250), (300, 300), (400, 400)\}$, and set the other parameters as the same as the Poisson and Gamma regression models described in §S6.1. The simulation results of the logistic regression model with crossed random effects based on 500 Monte Carlo studies are reported in Figures S4, S5, and S6. We found that the computational time of the proposed method is much smaller than for the Laplace approximation in the package `glmmTMB` and a GVA appled to the full log-likelihood function.

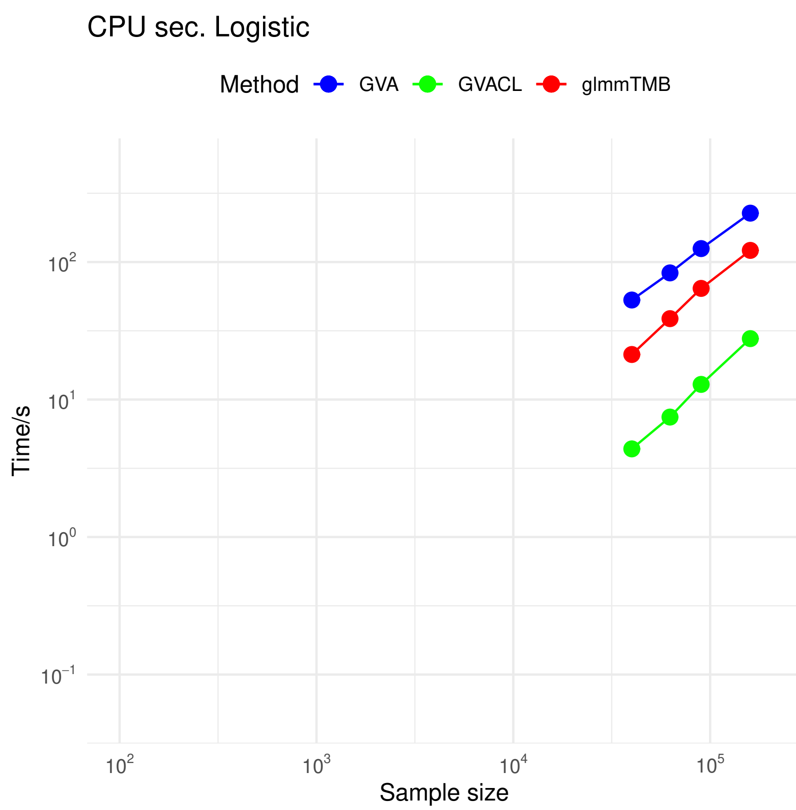

Figure S4: Average time of each simulation in seconds versus sample size $m * n$ under four different scenarios for the logistic regression model with crossed random effects. The ratio $\log m / \log n$ converges to 1.

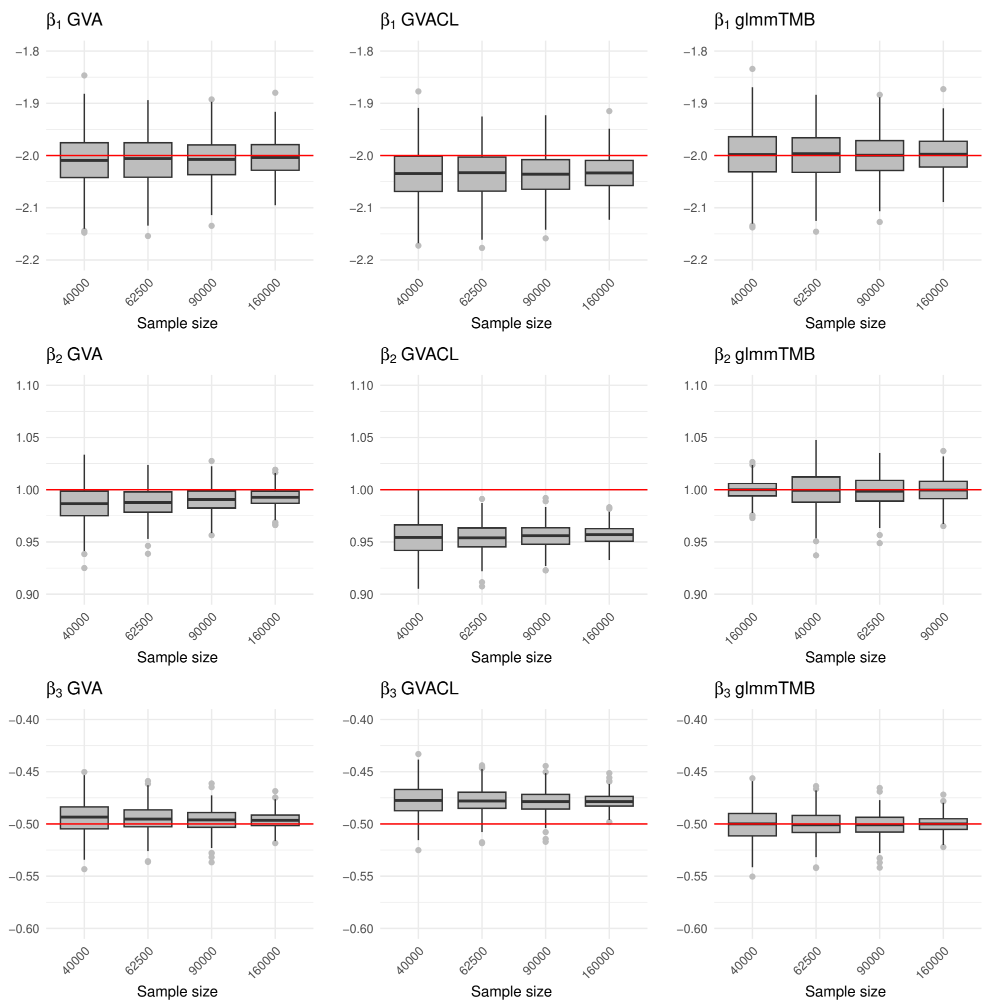

Figure S5: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA, GVACL, and `glmmTMB` for the logistic regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m / \log n$ converges to 1.

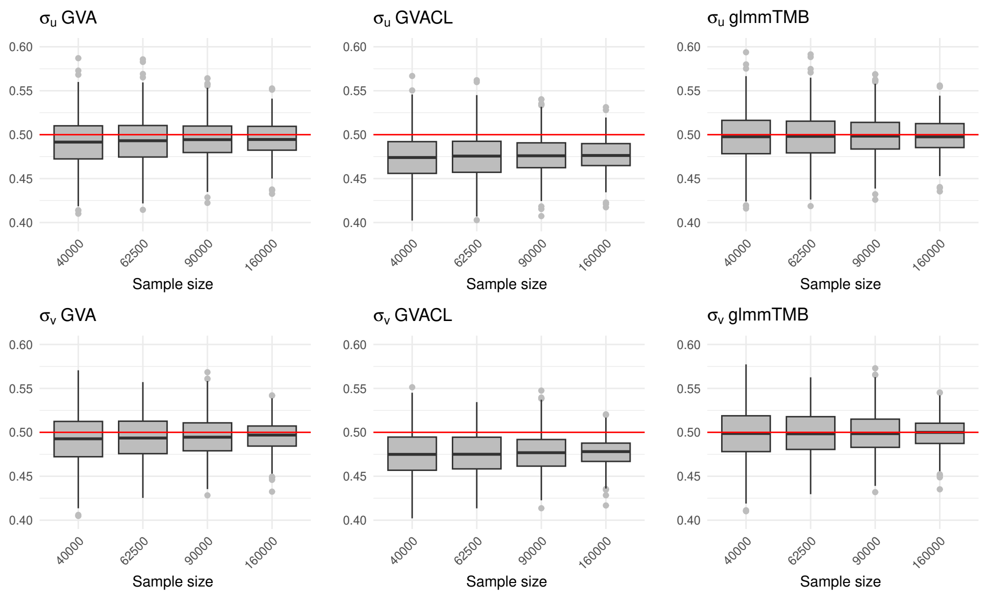

Figure S6: Boxplots of σ_u and σ_v estimates obtained by GVA, GVACL, and `glmmTMB` for the logistic regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m / \log n$ converges to 1.

S6.3 Simulations for Poisson, Gamma, and logistic regression models with $\delta = 1.1$

We also conduct the simulation studies to assess the performance of the Poisson, Gamma, and logistic regression models with crossed random effects. Except the case in the manuscript, we also consider another case, independently generate $\mathbf{X}_{ij,-1}$ from $N(0, I_{8\times 8})$ for $i = 1, ..., m$ and $j = 1, ..., n$, where $I_{8\times 8}$ is an identity matrix. We set $\boldsymbol{\beta}^0 = (-2, 1, 0.5, 1, -1, 0.5, 0.8, -0.4, 0.1)$, $(\sigma_u)^0 = 0.5$, and $(\sigma_v)^0 = 0.5$. The additional shape parameter is set as $\alpha = 0.8$. For Poisson and Gamma cases, we consider nine different scenarios : $(m, n) \in \{(18, 14), (20, 15), (30, 22), (50, 35), (80, 54), (100, 66), (150, 95), (200, 124), (300, 179)\}$, which imply the nine increasing sample sizes $m * n \in \{252, 300, 660, 1750, 4320, 6600, 14250, 24800, 53700\}$. Most Monte Carlo simulations of GVA to logistic case fails to converge when the sample size is small. Hence we consider four different scenarios: $(m, n) \in \{(200, 124), (250, 151), (300, 179), (400, 232)\}$. We report the results based on 500 Monte Carlo studies in Figures S8-S15.

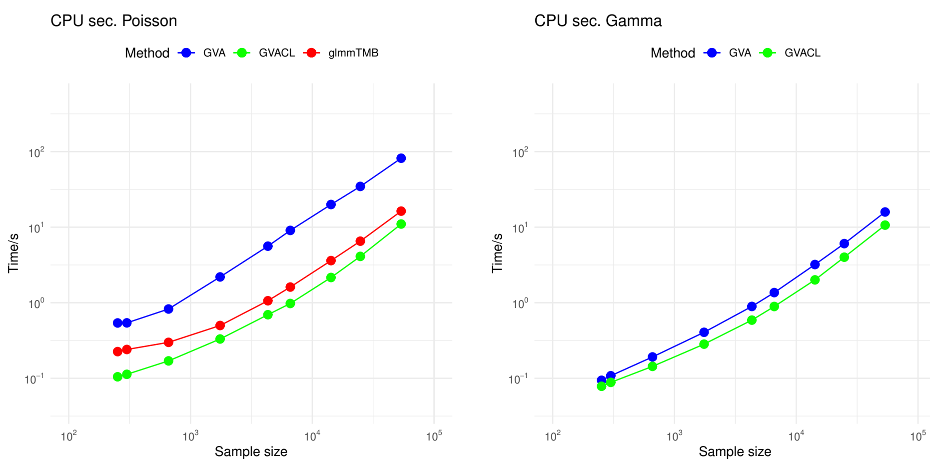

Figure S7: Average time of each simulation in seconds versus sample size $m * n$ under different scenarios for the Poisson and Gamma regression models with crossed random effects. The ratio $\log m / \log n$ converges to 1.1.

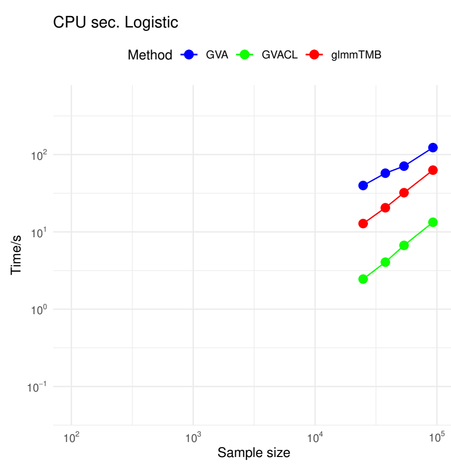

Figure S8: Average time of each simulation in seconds versus sample size $m * n$ under different scenarios for the logistic regression models with crossed random effects. The ratio $\log m / \log n$ converges to 1.1.

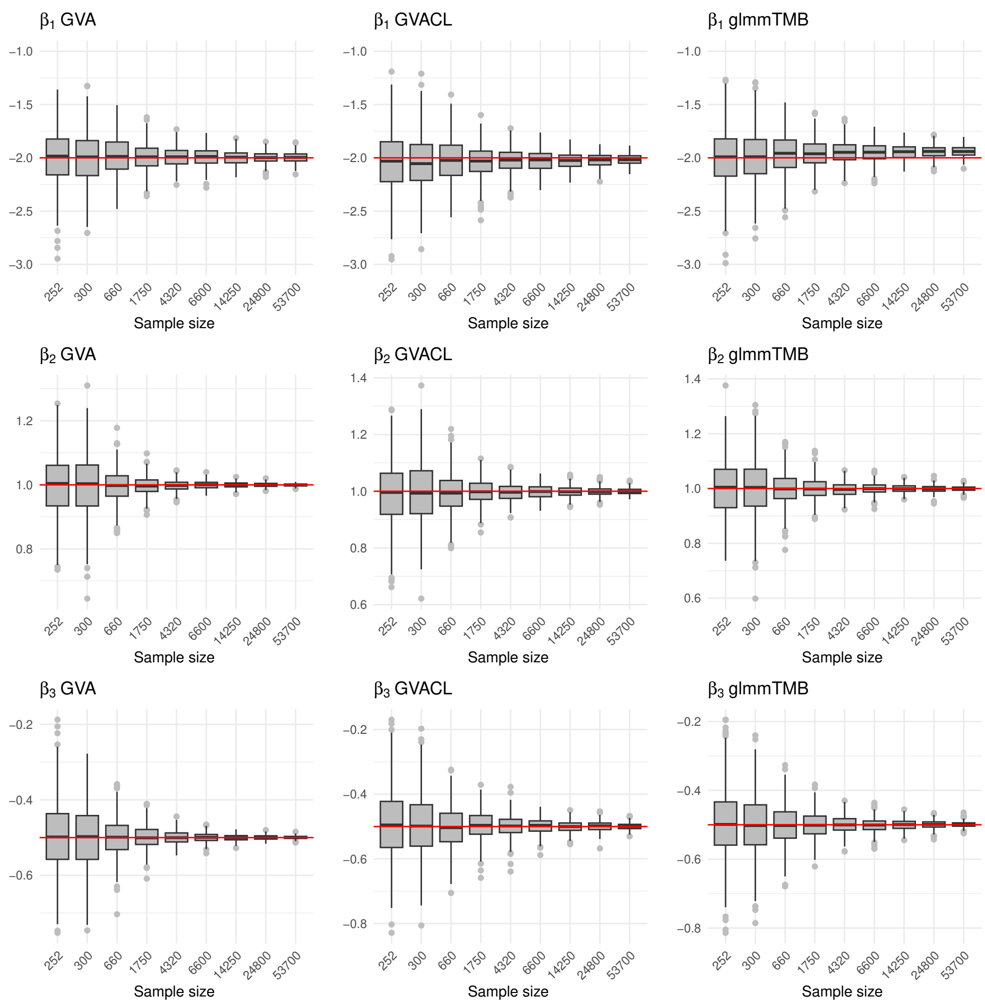

Figure S9: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA and GVACL for the Poisson regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m / \log n$ converges to 1.1.

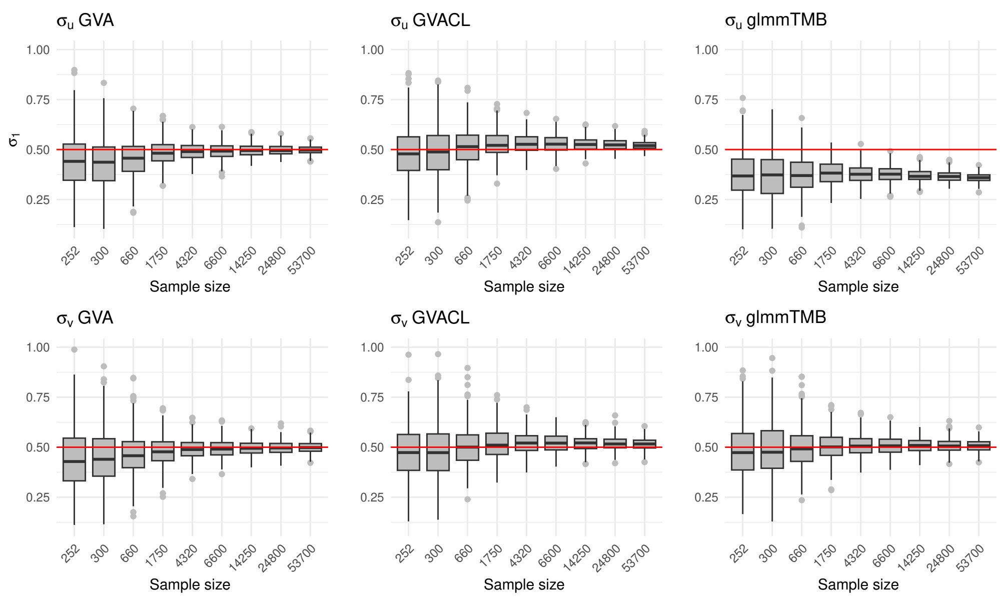

Figure S10: Boxplots of σ_u and σ_v estimates obtained by GVA and GVACL for the Poisson regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.1.

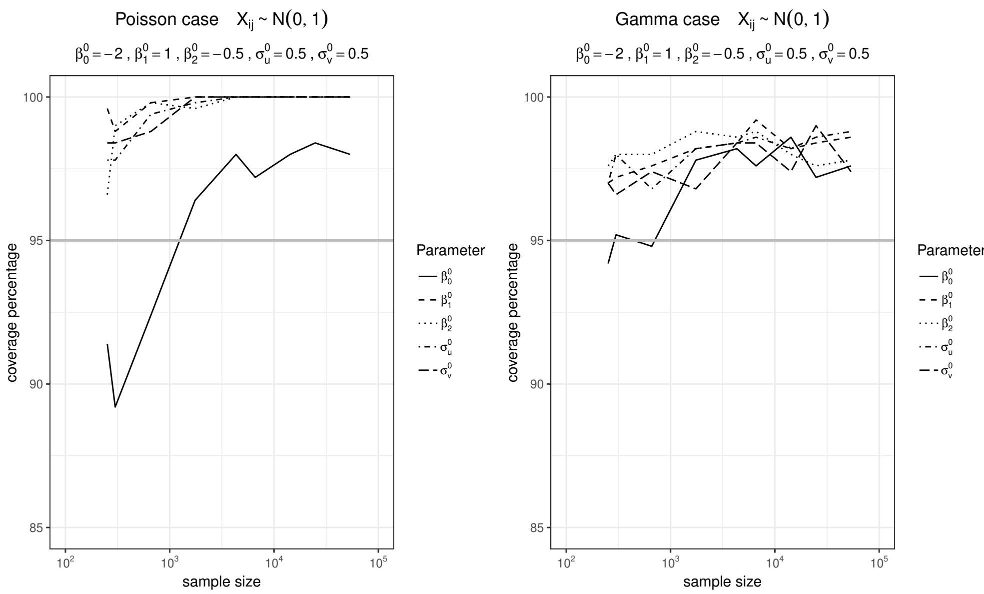

Figure S11: The coverage percentage of the 95% GVACL confidence intervals for the parameters in Poisson and Gamma regression models with crossed random effects based on 500 replications. The 95% percentage is shown as a thick grey horizontal line. The ratio $\log m / \log n$ converges to 1.1.

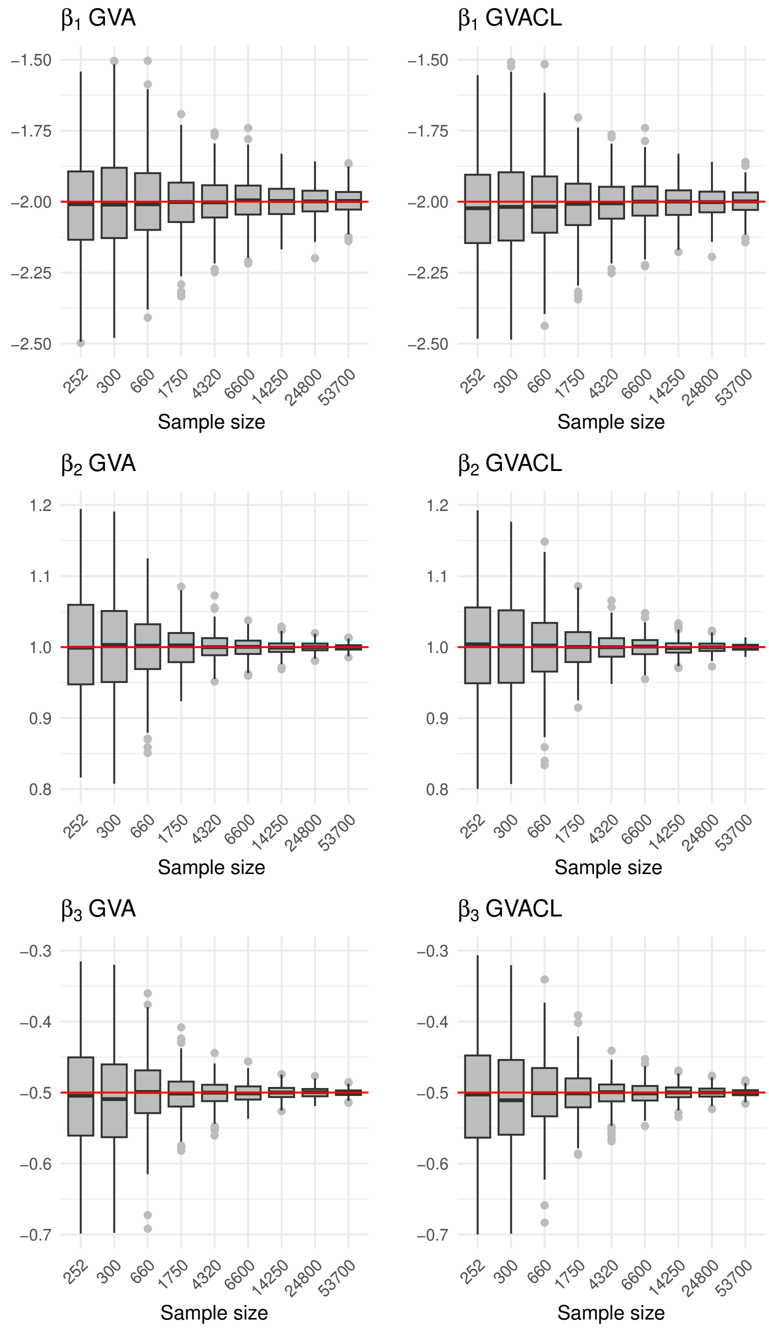

Figure S12: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA and GVACL for the Gamma regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.1.

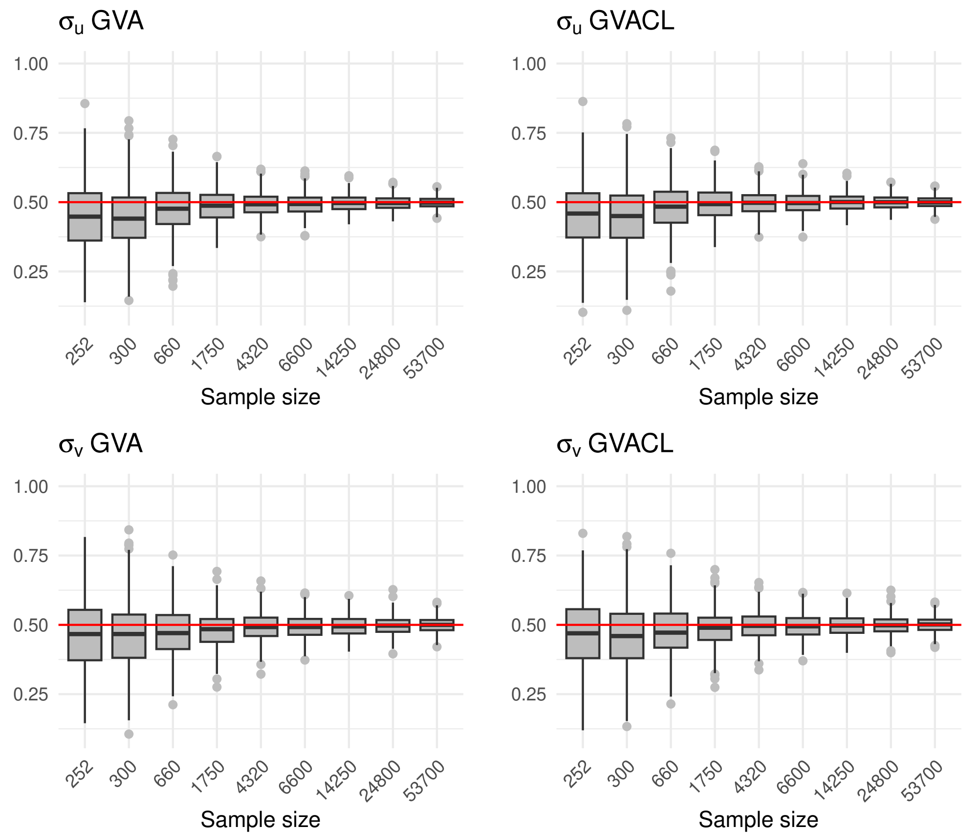

Figure S13: Boxplots of σ_u and σ_v estimates obtained by GVA and GVACL for the Gamma regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.1 .

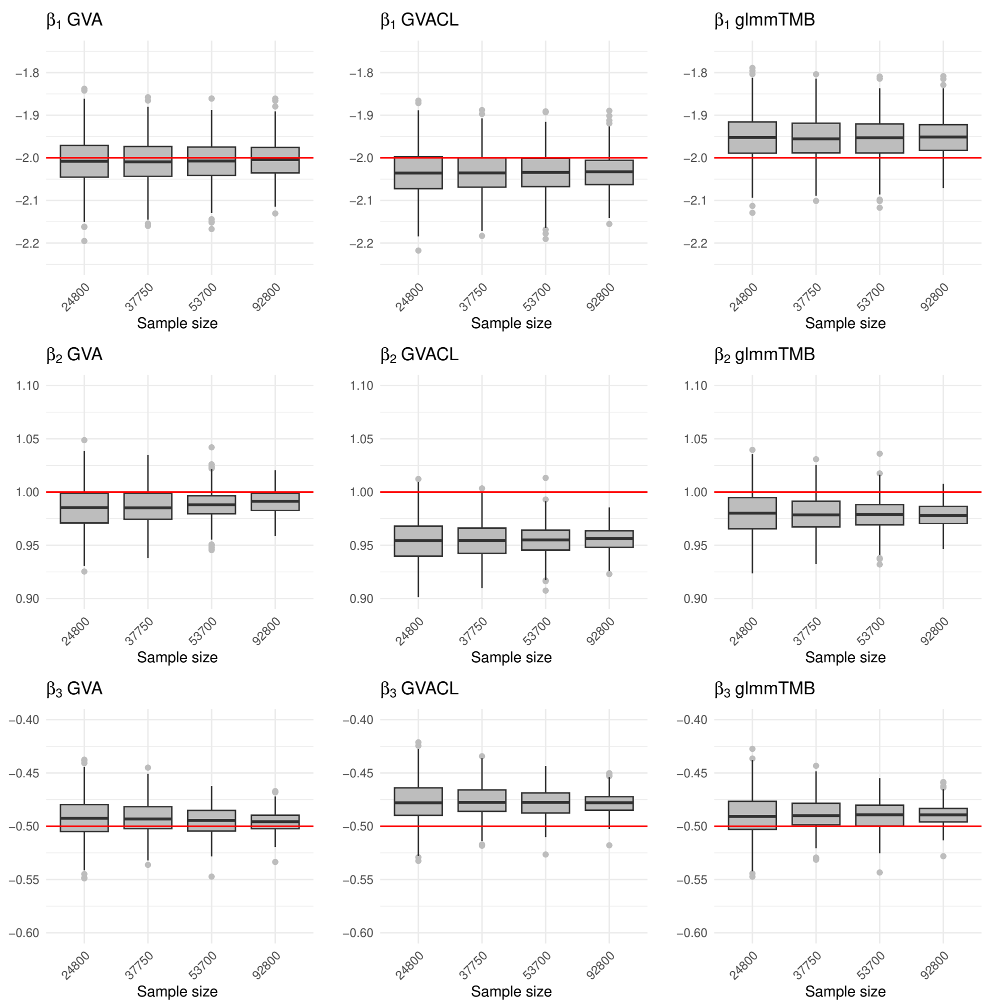

Figure S14: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA and GVACL for the logistic regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.1.

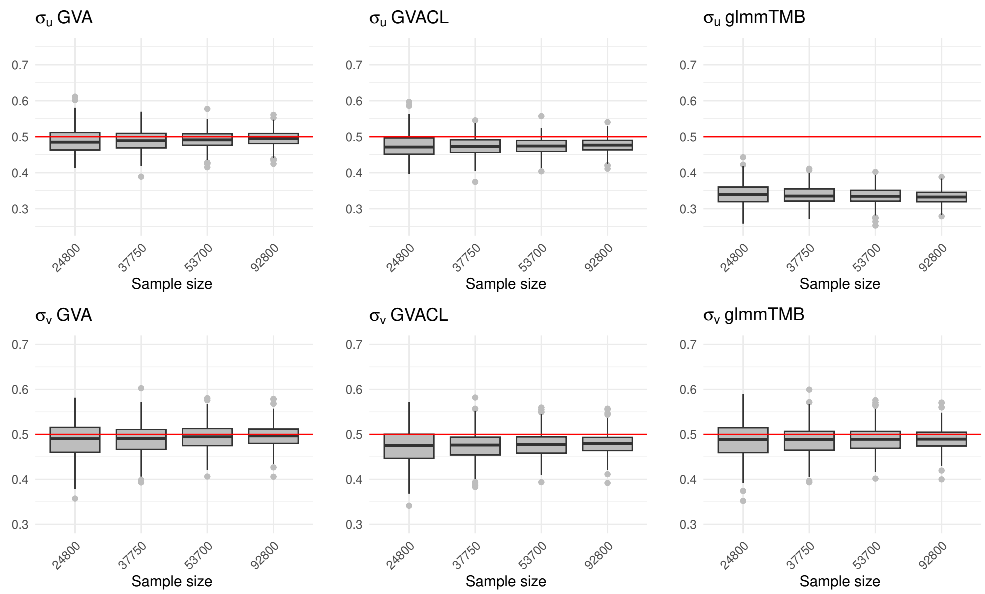

Figure S15: Boxplots of σ_u and σ_v estimates obtained by GVA and GVACL for the Gamma regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values. The ratio $\log m/\log n$ converges to 1.1 .

S7 Summary of the motor vehicle insurance data

In Table S1, we report the counts of claim events for each area and month,

Table S1: The counts of claim events for each area and month

Area code	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
01	5540	3388	4479	6840	7421	8068	5028	8841	6791	4982	1513	59
02	1100	610	811	1214	1338	1506	887	1443	1183	890	358	53
03	119	105	131	132	126	159	112	160	143	105	46	4
04	230	154	196	266	288	276	211	291	246	154	66	5
05	1240	734	1240	1580	1648	1853	1261	1853	1423	923	229	8
06	91	55	51	81	97	90	80	116	89	73	22	2
07	371	216	290	523	472	561	372	539	413	268	48	2
08	22	22	12	25	28	42	48	35	27	40	28	6
13	14	12	20	16	19	18	13	19	22	14	12	1
14	74	64	75	119	131	108	83	123	80	22	7	1
19	102	117	147	145	156	147	127	150	116	73	32	2
20	487	316	378	566	533	567	398	629	519	357	145	2
25	1347	818	976	1414	1590	1714	1170	1762	1306	963	233	20
26	251	143	203	352	364	420	243	470	316	193	50	1
27	515	311	454	711	827	902	594	973	701	323	106	2
31	179	115	151	212	227	299	179	347	245	122	56	1
33	519	327	436	557	605	676	412	723	583	410	122	9
34	453	210	381	589	679	706	449	835	645	446	120	14

We randomly selected one claim event from each area and month. A portion of the dataset can be found in Table S2.

Table S2: 18×12 samples are randomly selected from 18 areas and 12 months. The deductible and the amount of claim are both in units of 10 million rupiahs.

Month of occurence	Area code	Deductible	Amount of claim
1	1	0.13	0.36
1	2	0.02	0.21
1	3	0.12	0.33
1	4	0.02	0.10
1	5	0.02	0.13
1	6	0.02	0.39
1	7	0.04	0.38
1	8	0.16	0.35
1	13	0.02	0.17
1	14	0.02	0.02
1	19	0.02	1.27
1	20	0.02	0.05
1	25	0.04	0.08
1	26	0.02	0.06
1	27	0.04	0.11
1	31	0.02	1.53
1	33	0.08	0.88
1	34	0.06	0.08
2	1	0.02	1.32
2	2	0.02	1.29

Bibliography

Hall, P., J. T. Ormerod, and M. P. Wand (2011). Theory of Gaussian variational approximation for a poisson mixed model. *Statistica Sinica 21*, 369–389.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics 40*(2), 633–643.

Ormerod, J. T. and M. P. Wand (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics 21*(1), 2–17.