Supplemental Materials of: fast two-stage estimator for clustered count data with overdispersion

ARTICLE HISTORY
Compiled January 9, 2019

Appendix A. Variance of $\tilde{\boldsymbol{D}}$

To find an expression for $V(\tilde{\boldsymbol{D}})$, the delta method is applied. First, defining:

$$h(\boldsymbol{D}) = \sum_{i=1}^{N}(\boldsymbol{I} - \boldsymbol{H}_{ii})V\left(\hat{\boldsymbol{\beta}}_i\right)(\boldsymbol{I} - \boldsymbol{H}_{ii})' + \sum_{k\neq i}\boldsymbol{H}_{ik}V\left(\hat{\boldsymbol{\beta}}_i\right)\boldsymbol{H}'_{ik}, \tag{A1}$$

the variance of $\tilde{\boldsymbol{D}}$ equals:

$$V\left(\tilde{\boldsymbol{D}}\right) = \left[h'(\tilde{\boldsymbol{D}})^{-1}\right]' V\left(\text{vec}\ \ \boldsymbol{S}_b\right)\left[h'(\tilde{\boldsymbol{D}})^{-1}\right],$$

where $h'(\tilde{\boldsymbol{D}})$ is the derivative of $h(\boldsymbol{D})$ with respect to $\boldsymbol{D}$ evaluated at $\tilde{\boldsymbol{D}}$ and vec is the vec-operator. vec $\boldsymbol{A}$ vectorizes a matrix $\boldsymbol{A}$ by stacking its columns.

Assuming that asymptotically $\tilde{\boldsymbol{b}}_i$ follows a normal distribution, and using some results of Theorem 4.3 by [1], the variance of vec $\boldsymbol{S}_b$ is:

$$\begin{aligned}
V\left(\text{vec}\ \ \boldsymbol{S}_b\right) =& V\left(\sum_{i=1}^{N}\text{vec}\ \ \tilde{\boldsymbol{b}}_i\tilde{\boldsymbol{b}}_i'\right) = \sum_{i=1}^{N}V(\tilde{\boldsymbol{b}}_i\tilde{\boldsymbol{b}}_i') + \sum_{j\neq i}\text{Cov}(\tilde{\boldsymbol{b}}_i\tilde{\boldsymbol{b}}_i', \tilde{\boldsymbol{b}}_j\tilde{\boldsymbol{b}}_j') \\
=& \sum_{i=1}^{N}V\left(\tilde{\boldsymbol{b}}_i^{\otimes 2}\right) + \sum_{j\neq i}\text{Cov}\left(\tilde{\boldsymbol{b}}_i^{\otimes 2}, \tilde{\boldsymbol{b}}_j^{\otimes 2}\right) \qquad \text{(A2)} \\
=& (\boldsymbol{I} + \boldsymbol{C}_{qq})\left[\sum_{i=1}^{N}V\left(\tilde{\boldsymbol{b}}_i\right)^{\otimes 2} + 2\sum_{j>i}\text{Cov}\left(\tilde{\boldsymbol{b}}_i, \tilde{\boldsymbol{b}}_j\right)^{\otimes 2}\right],
\end{aligned}$$

where

$$\begin{aligned}
\text{Cov}\left(\tilde{\boldsymbol{b}}_i, \tilde{\boldsymbol{b}}_j\right) =& E\left(\tilde{\boldsymbol{b}}_i\tilde{\boldsymbol{b}}_j'\right) = E\left[\left(\hat{\boldsymbol{\beta}}_i - \boldsymbol{K}_i\tilde{\boldsymbol{\beta}}\right)\left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{K}_j\tilde{\boldsymbol{\beta}}\right)'\right] \\
=& \boldsymbol{K}_iV\left(\tilde{\boldsymbol{\beta}}\right)\boldsymbol{K}_j' - V\left(\hat{\boldsymbol{\beta}}_i\right)\boldsymbol{H}'_{ji} - \boldsymbol{H}_{ij}V\left(\hat{\boldsymbol{\beta}}_j\right),
\end{aligned}$$

and $\boldsymbol{C}_{pq}$ is the commutation matrix for an arbitrary $(p\times q)$ matrix $\boldsymbol{A}$, i.e., the matrix which, for any $(p\times q)$ matrix $\boldsymbol{A}$, transforms vec $\boldsymbol{A}$ into vec $\boldsymbol{A}'$.

Appendix B. Random intercept model

Here, the model is expressed as:

$$\boldsymbol{Y}_i|b_i \sim \text{Poisson}(\boldsymbol{\mu}_i^c), \tag{B1}$$

where $\boldsymbol{\mu}_i^c = \exp\left(\boldsymbol{X}_i\boldsymbol{\beta} + \mathbf{1}_{n_i}b_i\right)$ and $b_i \sim N(0,d)$. Then, the conditional variance of $\boldsymbol{Y}_i$ is reduced to,

$$V(\boldsymbol{Y}_i|b_i) = \exp(b_i)\boldsymbol{R}_i, \tag{B2}$$

where $\boldsymbol{R}_i$ is a $(n_i \times n_i)$ diagonal matrix with the vector $\exp\left(\boldsymbol{X}_i\boldsymbol{\beta}\right)$ at the diagonal.

Using the IRLS estimator for $\boldsymbol{\beta}_i$, the conditional mean and variance of $\hat{\boldsymbol{\beta}}_i$ are

$$E(\hat{\boldsymbol{\beta}}_i|b_i) = \boldsymbol{K}_i\boldsymbol{\beta} + b_i\boldsymbol{c} \quad \text{and} \quad V(\hat{\boldsymbol{\beta}}_i|b_i) = \exp\left(-b_i\right)\left(\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{T}_i\right)^{-1}, \tag{B3}$$

where $\boldsymbol{c}$ is a q-dimensional single-entry vector with first entry equals to 1 and zero elsewhere. Then, the marginal variance of $\hat{\boldsymbol{\beta}}_i$ is reduced to,

$$V\left(\hat{\boldsymbol{\beta}}_i\right) = d\boldsymbol{C} + \exp\left(\frac{d}{2}\right)\left(\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{T}_i\right)^{-1}, \tag{B4}$$

where $\boldsymbol{C}$ is a $(q\times q)$ single-entry matrix with entry $(1,1)$ equals to 1 and zero elsewhere. The variance of $\tilde{\boldsymbol{b}}$ is,

$$\begin{aligned}
V(\tilde{\boldsymbol{b}}_i) =& (\boldsymbol{I}-\boldsymbol{H}_{ii})V(\hat{\boldsymbol{b}}_i)(\boldsymbol{I}-\boldsymbol{H}_{ii})' + \sum_{k\neq i}\boldsymbol{H}_{ik}V(\hat{\boldsymbol{\beta}}_k)\boldsymbol{H}_{ik}' \\
=& d\left[(\boldsymbol{I}-\boldsymbol{H}_{ii})\boldsymbol{C}(\boldsymbol{I}-\boldsymbol{H}_{ii})' + \sum_{k\neq i}\boldsymbol{H}_{ik}\boldsymbol{C}\boldsymbol{H}_{ik}'\right] + \\
& \exp\left(\frac{d}{2}\right)\left[(\boldsymbol{I}-\boldsymbol{H}_{ii})\left(\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{T}_i\right)^{-1}(\boldsymbol{I}-\boldsymbol{H}_{ii})' + \sum_{k\neq i}\boldsymbol{H}_{ik}\left(\boldsymbol{T}_i'\boldsymbol{R}_i\boldsymbol{T}_i\right)^{-1}\boldsymbol{H}_{ik}'\right].
\end{aligned} \tag{B5}$$

Then, the estimator of d is based on:

$$s_b = \sum_{i=1}^{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{K}_i\tilde{\boldsymbol{\beta}})'(\hat{\boldsymbol{\beta}} - \boldsymbol{K}_i\tilde{\boldsymbol{\beta}}) = \sum_{i=1}^{N}\tilde{\boldsymbol{b}}_i'\tilde{\boldsymbol{b}}_i. \tag{B6}$$

A unbiased method-of-moments estimator of d can be found by equating s_b with its expected value and solving for d. Given that $E(\tilde{\boldsymbol{b}}_i) = \mathbf{0}$,

$$E(s_b) = \sum_{i=1}^{N} \operatorname{tr} V\left(\tilde{\boldsymbol{b}}_i\right) = d \sum_{i=1}^{N} \operatorname{tr} \left[(\boldsymbol{I} - \boldsymbol{H}_{ii}) \boldsymbol{C} (\boldsymbol{I} - \boldsymbol{H}_{ii})' + \sum_{k \neq i} \boldsymbol{H}_{ik} \boldsymbol{C} \boldsymbol{H}'_{ik} \right] + \exp\left(\frac{d}{2}\right) \sum_{i=1}^{N} \operatorname{tr} \left[(\boldsymbol{I} - \boldsymbol{H}_{ii}) \boldsymbol{C}_{2i} (\boldsymbol{I} - \boldsymbol{H}_{ii})' + \sum_{k \neq i} \boldsymbol{H}_{ik} \boldsymbol{C}_{2k} \boldsymbol{H}'_{ik} \right]. \tag{B7}$$

As before (B7) is non-linear and an iterative procedure is needed to find the solution for d.

Furthermore, assuming that $\tilde{b}_i$ follows a normal distribution, the variance of s_b is:

$$V(s_b) = \sum_{i=1}^{N} \operatorname{tr} \left[V\left(\tilde{\boldsymbol{b}}_i\right) V\left(\tilde{\boldsymbol{b}}_i\right) \right]. \tag{B8}$$

Then, an expression of $V\left(\tilde{d}\right)$ can be found using the delta method, as in Section A.

Appendix C. More results of the simulation study

In this section, we compare the cluster-by-cluster estimator of the Poisson-Normal model with different weighting schemes and the coverage is evaluated. For the Poisson-Normal-Gamma model, the results are fairly the same.

Weighting scheme comparison

Table C1 displays the MSE ratio of the cluster-by-cluster estimator using proportional over iterated optimal weights. For all parameters, the use of the latter reduces the MSE, but by a small (around 5% for fixed effects) or insignificant (less than 1% for variance components) quantity.

Table C1. MSE ratio of the cluster-by-cluster estimator using proportional weights over iterated optimal weights.

	(a) Fixing $N = 50$								(b) Fixing $\mu_n = 50$						
μ_n	β_0	β_1	β_2	β_3	d_{00}	d_{01}	d_{11}	N	β_0	β_1	β_2	β_3	d_{00}	d_{01}	d_{11}
20	1.02	1.04	1.03	1.01	1.00	1.00	1.00	20	1.02	1.02	1.02	1.00	1.01	1.01	1.00
50	1.02	1.05	1.04	1.06	1.00	1.00	1.00	50	1.02	1.05	1.04	1.06	1.00	1.00	1.00
100	1.05	1.05	1.04	1.04	1.00	1.00	1.00	100	1.04	1.06	1.02	1.02	1.00	1.00	1.00
150	1.06	1.06	1.06	1.06	1.00	1.00	1.00	150	1.03	1.04	1.07	1.05	1.00	1.00	1.00
200	1.07	1.05	1.07	1.07	1.00	1.00	1.00	200	1.07	1.06	1.03	1.05	1.00	1.00	1.00
250	1.07	1.04	1.06	1.05	1.00	1.00	1.00	250	1.05	1.05	1.05	1.05	1.00	1.00	1.00
400	1.05	1.08	1.05	1.06	1.00	1.00	1.00	400	1.06	1.07	1.04	1.03	1.00	1.00	1.00

Table C2 exhibits the MSE ratio of the method-of-moments estimator of D without weighting over with proportional weights. For both models, the use of proportional weights reduces the MSE by roughly 14% when the number of elements per cluster is small ($\mu_n = 20$). However, when the number of elements per cluster increases the unweighted estimators is lightly more efficient (by around 5% when $\mu_n = 250$).

Table C2. MSE ratio of the method-of-moments estimator of $\boldsymbol{D}$ without any weights over proportional weights

	(a) Fixing $N = 50$				(b) Fixing $\mu_n = 50$		
μ_n	d_{00}	d_{01}	d_{11}	N	d_{00}	d_{01}	d_{11}
20	1.10	1.17	1.15	20	0.98	0.99	1.00
50	1.01	1.02	1.00	50	1.01	1.02	1.00
100	0.96	0.99	0.98	100	0.99	1.01	1.02
150	0.96	0.96	0.94	150	1.02	1.02	1.01
250	0.95	0.95	0.95	250	1.00	1.01	1.07

Coverage of the 95% confidence interval

Table C3 displays the coverage of the 95% confidence interval of the parameter associated with the treatment effect (β_3) using the CbC and full ML estimator for the PN and PNG model in all simulation scenarios. Both estimators provide fairly the same results, the coverage is close to 0.95 in all scenarios.

Table C3. Coverage of the 95% confidence interval of the parameter associated with the treatment effect using the cluster-by-cluster (CbC) and full maximum likelihood (ML) estimator for the (a) PN model and (b) PNG model

(a) Poisson-Normal model										
	Fixing $\mu_n = 50$ and varying N					Fixing $N = 50$ and varying μ_n				
Method	20	50	100	150	250	20	50	100	150	250
CbC	0.928	0.938	0.945	0.955	0.951	0.938	0.938	0.938	0.934	0.933
ML	0.932	0.941	0.947	0.953	0.951	0.942	0.941	0.941	0.933	0.935
	(b) Poisson-Normal-Gamma model									
	Fixing $\mu_n = 50$ and varying N					Fixing $N = 50$ and varying μ_n				
CbC	0.927	0.944	0.964	0.950	0.957	0.938	0.944	0.941	0.940	0.935
ML	0.936	0.948	0.965	0.951	0.955	0.935	0.948	0.944	0.943	0.938

References

[1] Magnus JR, Neudecker H. The commutation matrix: Some properties and applications. The Annals of Statistics. 1979;7(2):381–394.