Improved JLA Estimates for Calculation of Leave Out Bias Correction of Variance Components

Patrick Kline, Raffaele Saggio, Mikkel Sølvsten

January 5, 2021

This document describes computation of the terms $(P_{ii}, M_{ii}, \sigma_i^2)$ as defined in Kline, Saggio, and Sølvsten (2020) — KSS henceforth — using a variant of the JLA algorithm described in KSS.

There are two closely related ways to estimate $P_{ii} = x_i' S_{xx}^{-1} x_i$ and $M_{ii} = 1 - P_{ii}$ using the random projection algorithm introduced by Achlioptas (2003). Let $q_1, \ldots, q_p \in \mathbb{R}^n$ be independent Rademacher vectors with independent entries and construct

$$\hat{P}_{ii} = \frac{1}{p} \sum_{s=1}^{p} (x_i' S_{xx}^{-1} X' q_s)^2 \quad \text{and} \quad \hat{M}_{ii} = \frac{1}{p} \sum_{s=1}^{p} (q_{si} - x_i' S_{xx}^{-1} X' q_s)^2$$

which are unbiased estimators of P_{ii} and M_{ii}, respectively. The covariance structure of these two estimators is

$$V(\hat{P}_{ii}) = \frac{2}{p} \left(P_{ii}^2 - \sum_{j=1}^{n} P_{ij}^4 \right), \; V(\hat{M}_{ii}) = \frac{2}{p} \left(M_{ii}^2 - \sum_{j=1}^{n} M_{ij}^4 \right),$$
$$C(\hat{P}_{ii}, \hat{M}_{ii}) = \frac{2}{p} \left(0 - \sum_{j=1}^{n} P_{ij}^2 M_{ij}^2 \right).$$

When estimating M_{ii}, the infeasible variance minimizing unbiased linear combination of these two estimators is

$$\frac{P_{ii}}{M_{ii} + P_{ii}} \hat{M}_{ii} + \frac{M_{ii}}{M_{ii} + P_{ii}} (1 - \hat{P}_{ii})$$

The feasible version that uses hats everywhere takes a very simple form and is give by

$$\bar{M}_{ii} = \frac{\hat{M}_{ii}}{\hat{M}_{ii} + \hat{P}_{ii}} \tag{1}$$

Remark 1. Note that because $\hat{M}_{ii}$ and $\hat{P}_{ii}$ are both non-negative, we get an estimate that is guaranteed to lie inside $[0, 1]$. The corresponding estimator of P_{ii} is $\bar{P}_{ii} = \frac{\hat{P}_{ii}}{\hat{M}_{ii}+\hat{P}_{ii}}$, so we are simply imposing on the estimators that they satisfy the constraint that $M_{ii} + P_{ii} = 1$. Furthermore, the asymptotic variance (as p gets large) of the constrained estimator $\bar{M}_{ii}$ is

$$V_i = \frac{2}{p}\left(2M_{ii}^2 P_{ii}^2 - (M_{ii} - P_{ii})^2 \sum_{j=1}^{n} P_{ij}^2 M_{ij}^2\right).$$

At the boundaries the new estimator has no variance improvement or loss relative to the best of $\hat{M}_{ii}$ and $\hat{P}_{ii}$, as it simply picks the best of the two. At the center of the support, $\hat{M}_{ii}$ and $\hat{P}_{ii}$ have equal variance and the new estimator improves on both as their correlation is different from -1. Finally, an inspection of the mean of $\bar{M}_{ii}$ reveals that it relies on shrinkage towards the middle of the support, i.e., the mean of $\bar{M}_{ii}$ (as p gets large) is

$$M_{ii} + B_i, \qquad B_i = \frac{2}{p}\left(P_{ii} - M_{ii}\right)\left(M_{ii}P_{ii} + 2\sum_{j=1}^{n} P_{ij}^2 M_{ij}^2\right).$$

The estimator highlighted in (1) avoids non-sensical leverage estimates outside of the support, and it reduces variance (thus it allows for fewer repetitions, p, in the JLA algorithm). Furthermore, it does not require extra matrix inversions, as both $\hat{M}_{ii}$ and $\hat{P}_{ii}$ can be constructed after one call to pcg per Rademacher vector.

Non-linearity bias

The JLA approximation of $\hat{\sigma}_i^2 = \dfrac{y_i(y_i - x_i'\hat{\beta})}{M_{ii}}$ is given by $\tilde{\sigma}_i^2 = \dfrac{y_i(y_i - x_i'\hat{\beta})}{\bar{M}_{ii}}$. One can see that $\tilde{\sigma}_i^2$ has a mean (conditional on data) approximated to second order that is

$$\hat{\sigma}_i^2\left(1 + \frac{V_i}{M_{ii}^2} - \frac{B_i}{M_{ii}}\right).$$

We can construct an estimator of the bias using the same Rademacher draws that we use to construct $\hat{M}_{ii}$ and $\hat{P}_{ii}$ and define therefore the following estimator

$$\hat{\sigma}_{i,JLA}^2 = \frac{y_i(y_i - x_i'\hat{\beta})}{\bar{M}_{ii}}\left(1 - \frac{\hat{V}_i}{\bar{M}_{ii}^2} + \frac{\hat{B}_i}{\bar{M}_{ii}}\right).$$

The main advantage of this correction relative to the original one proposed in KSS is that it removes the entire bias of order p^{-1} in the JLA point estimator.[1] We suspect that the next order bias is then of order p^{-2}, thus there is no bias of importance as long as $n/p^4 = o(1)$.

Define the following three second moment estimators

$$m(P_{ii}^2) = \frac{1}{p}\sum_{s=1}^{p}(x_i' S_{xx}^{-1} X' q_s)^4, \; m(M_{ii}^2) = \frac{1}{p}\sum_{s=1}^{p}(q_{si} - x_i' S_{xx}^{-1} X' q_s)^4$$
$$m(P_{ii}, M_{ii}) = \frac{1}{p}\sum_{s=1}^{p}(x_i' S_{xx}^{-1} X' q_s)^2 (q_{si} - x_i' S_{xx}^{-1} X' q_s)^2.$$

We can then define

$$\hat{V}_i = \frac{1}{p}\left(\bar{M}_{ii}^2 m(P_{ii}^2) + \bar{P}_{ii}^2 m(M_{ii}^2) - 2\bar{P}_{ii}\bar{M}_{ii} m(P_{ii}, M_{ii})\right)$$
$$\hat{B}_i = \frac{1}{p}\left(\bar{M}_{ii} m(P_{ii}^2) - \bar{P}_{ii} m(M_{ii}^2) + (\bar{M}_{ii} - \bar{P}_{ii}) m(P_{ii}, M_{ii})\right).$$

References

Achlioptas, D. (2003). Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences 66*(4), 671–687.

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out estimation of variance components. *Econometrica 88*(5), 1859–1898.

[1] In our experience, based on various simulations and empirical applications, we found that this non-linear bias to show up in small-scale applications where both n and p were relatively small.