ESTIMATION AND INFERENCE FOR VERY LARGE LINEAR MIXED EFFECTS MODELS

Katelyn Gao and Art B. Owen

Intel Inc. and Stanford University

Supplementary Material

This supplementary document contains proofs for results in “Estimating and Inference for Very Large Linear Mixed Effects Models” by Katelyn Gao and Art B. Owen.

S1 Proofs for Section 2

First, we repeat the exact variance formulas for U statistics used in [2]. For $\eta_{ij} = a_i + b_j + e_{ij}$, let

$$\begin{aligned} U_a = U_a(\beta) &= \frac{1}{2}\sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'}(\eta_{ij} - \eta_{ij'})^2, \\ U_b = U_b(\beta) &= \frac{1}{2}\sum_{jii'} N_{\bullet j}^{-1} Z_{ij} Z_{i'j}(\eta_{ij} - \eta_{i'j})^2, \quad \text{and} \\ U_e = U_e(\beta) &= \frac{1}{2}\sum_{iji'j'} Z_{ij} Z_{i'j'}(\eta_{ij} - \eta_{i'j'})^2. \end{aligned} \tag{S1.1}$$

The model from [2] applied those U-statistics to Y_{ij} instead of η_{ij}. In our notation, their Y_{ij} is $\mu + \eta_{ij}$. Because the intercept μ cancels, these U-statistics defined via η_{ij} are equivalent to those defined via Y_{ij}.

Theorem 1. *Let Y_{ij} follow the random effects model* (1.1) *with the observation pattern Z_{ij} as described in Section 2. Then the U-statistics defined at* (S1.1) *have variances*

$$\begin{aligned} \mathrm{Var}(U_a) = {} & \sigma_B^4(\kappa_B + 2)\sum_{ir}(ZZ^\mathsf{T})_{ir}(1 - N_{i\bullet}^{-1})(1 - N_{r\bullet}^{-1}) \\ & + 2\sigma_B^4 \sum_{ir} N_{i\bullet}^{-1} N_{r\bullet}^{-1} (ZZ^\mathsf{T})_{ir}[(ZZ^\mathsf{T})_{ir} - 1] + 4\sigma_B^2\sigma_E^2(N - R) \\ & + \sigma_E^4(\kappa_E + 2)\sum_i N_{i\bullet}(1 - N_{i\bullet}^{-1})^2 + 2\sigma_E^4 \sum_i (1 - N_{i\bullet}^{-1}), \end{aligned} \tag{S1.2}$$

and

$$\begin{aligned}\mathrm{Var}(U_b) = \sigma_A^4(\kappa_A+2)\sum_{js}(Z^{\mathsf{T}}Z)_{js}(1-N_{\bullet j}^{-1})(1-N_{\bullet s}^{-1}) \\ + 2\sigma_A^4\sum_{js}N_{\bullet j}^{-1}N_{\bullet s}^{-1}(Z^{\mathsf{T}}Z)_{js}[(Z^{\mathsf{T}}Z)_{js}-1] + 4\sigma_A^2\sigma_E^2(N-C) \\ + \sigma_E^4(\kappa_E+2)\sum_j N_{\bullet j}(1-N_{\bullet j}^{-1})^2 + 2\sigma_E^4\sum_j(1-N_{\bullet j}^{-1}),\end{aligned} \tag{S1.3}$$

and

$$\begin{aligned}\mathrm{Var}(U_e) = 2\sigma_A^4\Big[\Big(\textstyle\sum_i N_{i\bullet}^2\Big)^2 - \textstyle\sum_i N_{i\bullet}^4\Big] + 2\sigma_B^4\Big[\Big(\textstyle\sum_j N_{\bullet j}^2\Big)^2 - \displaystyle\sum_j N_{\bullet j}^4\Big] \\ + \sigma_A^4(\kappa_A+2)\Big(N^2\sum_i N_{i\bullet}^2 - 2N\sum_i N_{i\bullet}^3 + \sum_i N_{i\bullet}^4\Big) \\ + \sigma_B^4(\kappa_B+2)\Big(N^2\sum_j N_{\bullet j}^2 - 2N\sum_j N_{\bullet j}^3 + \sum_j N_{\bullet j}^4\Big) \\ + 2\sigma_E^4 N(N-1) + \sigma_E^4(\kappa_E+2)N(N-1)^2 \\ + 4\sigma_A^2\sigma_B^2\Big(N^3 - 2N\sum_{ij}Z_{ij}N_{i\bullet}N_{\bullet j} + \sum_{ij}N_{i\bullet}^2N_{\bullet j}^2\Big) \\ + 4\sigma_A^2\sigma_E^2\Big(N^3 - N\sum_i N_{i\bullet}^2\Big) + 4\sigma_B^2\sigma_E^2\Big(N^3 - N\sum_j N_{\bullet j}^2\Big).\end{aligned} \tag{S1.4}$$

Proof. This is a portion of [2, Theorem 4.1]. The remainder of that theorem gives the covariances among the three U-statistics. □

S1.1 Proof of Theorem 1

Proof. Letting $\epsilon = \max(\epsilon_R, \epsilon_C)$, we have

$$M = \begin{pmatrix} N & & \\ & N & \\ & & N^2 \end{pmatrix} \begin{pmatrix} 0 & 1-R/N & 1-R/N \\ 1-C/N & 0 & 1-C/N \\ 1 & 1 & 1 \end{pmatrix} (1+O(\epsilon))$$

and so if $\max(R,C)/N \leqslant \theta$ for some $\theta < 1$, then

$$M^{-1} = \begin{pmatrix} -\frac{N}{N-R} & 0 & 1 \\ 0 & -\frac{N}{N-C} & 1 \\ \frac{N}{N-R} & \frac{N}{N-C} & -1 \end{pmatrix} \begin{pmatrix} N^{-1} & & \\ & N^{-1} & \\ & & N^{-2} \end{pmatrix} (1+O(\epsilon)).$$

It follows that

$$\begin{aligned} \hat\sigma^2_A &= \Bigl(\frac{U_e}{N^2} - \frac{U_a}{N-R}\Bigr)(1+O(\epsilon)) \\ \hat\sigma^2_B &= \Bigl(\frac{U_e}{N^2} - \frac{U_b}{N-C}\Bigr)(1+O(\epsilon)), \quad \text{and} \\ \hat\sigma^2_E &= \Bigl(\frac{U_a}{N-R} + \frac{U_b}{N-C} - \frac{U_e}{N^2}\Bigr)(1+O(\epsilon)). \end{aligned} \tag{S1.5}$$

Gao and Owen [2, Lemma 4.1] show that $\mathbb{E}(U_a) = (\sigma^2_B + \sigma^2_E)(N-R)$, $\mathbb{E}(U_b) = (\sigma^2_A + \sigma^2_E)(N-C)$, and

$$\begin{aligned} \mathbb{E}(U_e) &= \sigma^2_A\Bigl(N^2 - \sum_i N^2_{i\bullet}\Bigr) + \sigma^2_B\Bigl(N^2 - \sum_j N^2_{\bullet j}\Bigr) + \sigma^2_E(N^2-N), \quad \text{so} \\ \frac{\mathbb{E}(U_e)}{N^2} &= \sigma^2_A + \sigma^2_B + \sigma^2_E - \Upsilon, \quad \text{where} \\ \Upsilon &= \Bigl(\sigma^2_A \frac{\sum_i N^2_{i\bullet}}{N^2} + \sigma^2_B \frac{\sum_j N^2_{\bullet j}}{N^2} + \frac{\sigma^2_E}{N}\Bigr) = O(\epsilon). \end{aligned}$$

By substitution in (S1.5) we find that all of the variance component biases are $\Upsilon \times (1+O(\epsilon)) = O(\epsilon)$.

Turning now to variances,

$$\begin{aligned} \mathrm{Var}(\hat\sigma^2_A) &= O\Bigl(\frac{\mathrm{Var}(U_e)}{N^4} + \frac{\mathrm{Var}(U_a)}{N^2}\Bigr), \\ \mathrm{Var}(\hat\sigma^2_B) &= O\Bigl(\frac{\mathrm{Var}(U_e)}{N^4} + \frac{\mathrm{Var}(U_b)}{N^2}\Bigr), \quad \text{and} \\ \mathrm{Var}(\hat\sigma^2_E) &= O\Bigl(\frac{\mathrm{Var}(U_a)}{N^2} + \frac{\mathrm{Var}(U_b)}{N^2} + \frac{\mathrm{Var}(U_e)}{N^4}\Bigr). \end{aligned} \tag{S1.6}$$

Some bounds for these variances are given by [2, Theorem 4.2]. That theorem makes stronger assumptions, such as a small bound on R/N that we do not want to make here, and so instead we work from the exact finite sample formulas in [2, Theorem 4.1], given here in Theorem 1. We note here that there is an error in [2, Section 9.8] where the coefficient of $\sigma_B^4(\kappa_B+2)$ in $\mathrm{Var}(U_a)$ is shown to be $\sum_j N_{\bullet j}^2(1+O(\delta))$ for the δ defined there. From the derivation there, it is clear that this coefficient is less than $\sum_j N_{\bullet j}^2$, and so the conclusion of that theorem is unaffected.

Using $(ZZ^\mathsf{T})_{ir} \leqslant N_{r\bullet}$ and $\sum_{ir}(ZZ^\mathsf{T})_{ir} = \sum_j N_{\bullet j}^2$, we find from (S1.2) that

$$
\begin{aligned}
\mathrm{Var}(U_a) &\leqslant \sigma_B^4(\kappa_B+2)\sum_{ir}(ZZ^\mathsf{T})_{ir} + 2\sigma_B^4\sum_{ir}N_{i\bullet}^{-1}(ZZ^\mathsf{T})_{ir} + 4\sigma_B^2\sigma_E^2 N \\
&\quad + \sigma_E^4(\kappa_E+2)\sum_i N_{i\bullet} + 2\sigma_E^4\sum_i 1 \\
&\leqslant \sigma_B^4(\kappa_B+4)\sum_j N_{\bullet j}^2 + \Bigl(4\sigma_B^2\sigma_E^2 + \sigma_E^4(\kappa_E+2)\Bigr)N + 2R\sigma_E^4 \\
&= O\Bigl(\sum_j N_{\bullet j}^2\Bigr). \qquad \text{(S1.7)}
\end{aligned}
$$

The same logic yields $\mathrm{Var}(U_b) = O(\sum_i N_{i\bullet}^2)$. The second term in $\mathrm{Var}(U_a)$, which was lumped in with the first, might ordinarily be much smaller than the first, and then a lead coefficient of $\sigma_B^4(\kappa_B+2)$ would be more accurate than $\sigma_B^4(\kappa_B+4)$.

For $\mathrm{Var}(U_e)$ the nonnegative terms in (S1.4) have magnitudes proportional to

$$\Big(\sum_i N_{i\bullet}^2\Big)^2,\ \Big(\sum_j N_{\bullet j}^2\Big)^2,\ N^2\sum_i N_{i\bullet}^2,\ N^2\sum_j N_{\bullet j}^2,$$
$$\sum_i N_{i\bullet}^4,\ \sum_j N_{\bullet j}^4,\ \sum_i N_{i\bullet}^2\sum_j N_{\bullet j}^2,\ N^3$$

or smaller. These are all $O(N^2(\sum_i N_{i\bullet}^2+\sum_j N_{\bullet j}^2))$, and so

$$\mathrm{Var}(U_e)=O\Big(N^2\Big(\sum_i N_{i\bullet}^2+\sum_j N_{\bullet j}^2\Big)\Big). \tag{S1.8}$$

Combining (S1.7) and (S1.8) into (S1.6) yields

$$\mathrm{Var}(\hat\sigma_A^2)=O\Big(\frac{\sum_i N_{i\bullet}^2}{N^2}+\frac{\sum_j N_{\bullet j}^2}{N^2}\Big)(1+O(\epsilon)),$$

and the same follows for $\hat\sigma_B^2$ by symmetry. Precisely the same terms appear in $\mathrm{Var}(\hat\sigma_E^2)$ so it also has that rate. □

S2 Proofs for Section 3

S2.1 Proof of Theorem 3

To compute a lower bound for $\mathrm{eff}_{\mathrm{RLS}}$, we first transform $\boldsymbol{x}$ into $\boldsymbol{z}=V_A^{-1/2}\boldsymbol{x}$. Then, from (3.8),

$$\mathrm{eff}_{\mathrm{RLS}}=\frac{(\boldsymbol{z}^T\boldsymbol{z})^2}{(\boldsymbol{z}^TV_A^{-1/2}V_RV_A^{-1/2}\boldsymbol{z})(\boldsymbol{z}^TV_A^{1/2}V_R^{-1}V_A^{1/2}\boldsymbol{z})}.$$

Scaling $\boldsymbol{z}$ by a nonzero constant does not change $\mathrm{eff}_{\mathrm{RLS}}$. Letting $\boldsymbol{u}=\boldsymbol{z}/\|\boldsymbol{z}\|$, we have

$$1/\mathrm{eff}_{\mathrm{RLS}}=(\boldsymbol{u}^TV_A^{-1/2}V_RV_A^{-1/2}\boldsymbol{u})(\boldsymbol{u}V_A^{1/2}V_R^{-1}V_A^{1/2}\boldsymbol{u})\equiv(\boldsymbol{u}^\mathsf{T}A\boldsymbol{u})(\boldsymbol{u}^\mathsf{T}A^{-1}\boldsymbol{u})$$

for $A = V_A^{-1/2} V_R V_A^{-1/2}$. We get an upper bound for $1/\mathrm{eff}_{\mathrm{RLS}}$ from the Kantorovich inequality after getting upper and lower bounds on the eigenvalues of

$$A = V_A^{-1/2}(V_A + \sigma_B^2 B_R) V_A^{-1/2} = I_N + \sigma_B^2 V_A^{-1/2} B_R V_A^{-1/2}.$$

The eigenvalues of A are the eigenvalues of $\sigma_B^2 V_A^{-1/2} B_R V_A^{-1/2}$ plus one.

The matrix B_C and by extension B_R is singular and positive semidefinite, with nonzero eigenvalues $N_{\bullet j}$ for $j = 1, \dots, C$. Also, V_A is symmetric and nonsingular with eigenvalues σ_E^2, and $\sigma_E^2 + \sigma_A^2 N_{i\bullet}$ for $i = 1, \dots, R$. Then $V_A^{-1/2}$ is symmetric and nonsingular with eigenvalues $1/\sqrt{\sigma_E^2}$ and $1/\sqrt{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}$ for $i = 1, \dots, R$.

Therefore, $\sigma_B^2 V_A^{-1/2} B_R V_A^{-1/2}$ is singular and positive semidefinite. Its smallest eigenvalue is zero, and its largest eigenvalue is bounded above by

$$\|\sigma_B^2 V_A^{-1/2} B_R V_A^{1/2}\|_2 \leqslant \sigma_B^2 \|V_A^{-1/2}\|_2^2 \|B_R\|_2 = \frac{\sigma_B^2}{\sigma_E^2} \max_j N_{\bullet j}.$$

This is where we needed the assumption that $\sigma_E^2 > 0$.

The smallest eigenvalue of A is 1 and the largest eigenvalue is at most $1 + \sigma_B^2 \max_j N_{\bullet j}/\sigma_E^2$. By the Kantorovich inequality (Theorem 2),

$$\begin{aligned}
1/\mathrm{eff}_{\mathrm{RLS}} &= (\boldsymbol{u}^T V_A^{-1/2} V_R V_A^{-1/2} \boldsymbol{u})(\boldsymbol{u}^T V_A^{1/2} V_R^{-1} V_A^{1/2} \boldsymbol{u}) \\
&\leqslant \frac{(2 + \sigma_B^2 \max_j N_{\bullet j}/\sigma_E^2)^2}{4(1 + \sigma_B^2 \max_j N_{\bullet j}/\sigma_E^2)} = \frac{(2\sigma_E^2 + \sigma_B^2 \max_j N_{\bullet j})^2}{4\sigma_E^2(\sigma_E^2 + \sigma_B^2 \max_j N_{\bullet j})}.
\end{aligned}$$

Taking reciprocals gives the desired result. The result for $\mathrm{eff}_{\mathrm{CLS}}$ follows by symmetry. The inequalities are tight because Kantorovich's inequality is tight.

S3 Proofs for Section 5

Several of the proofs for Section 5 utilize the following lemma, which is not given in the main paper for brevity's sake. This lemma rewrites $U_a(\hat\beta)$, $U_b(\hat\beta)$, and $U_e(\hat\beta)$ in a useful form.

Lemma 1.

$$
\begin{aligned}
U_a(\hat\beta) = U_a(\beta) &+ \frac{(\beta-\hat\beta)^{\mathsf{T}}}{2}\Bigl(\sum_{ijj'} N_{i\bullet}^{-1} Z_{ij}Z_{ij'}(x_{ij}-x_{ij'})(x_{ij}-x_{ij'})^{\mathsf{T}}\Bigr)(\beta-\hat\beta)\\
&+ (\beta-\hat\beta)^{\mathsf{T}}\sum_{ijj'} N_{i\bullet}^{-1} Z_{ij}Z_{ij'}(x_{ij}-x_{ij'})(b_j-b_{j'})\\
&+ (\beta-\hat\beta)^{\mathsf{T}}\sum_{ijj'} N_{i\bullet}^{-1} Z_{ij}Z_{ij'}(x_{ij}-x_{ij'})(e_{ij}-e_{ij'}),\\
U_b(\hat\beta) = U_b(\beta) &+ \frac{(\beta-\hat\beta)^{\mathsf{T}}}{2}\Bigl(\sum_{jii'} N_{\bullet j}^{-1} Z_{ij}Z_{i'j}(x_{ij}-x_{i'j})(x_{ij}-x_{i'j})^{\mathsf{T}}\Bigr)(\beta-\hat\beta)\\
&+ (\beta-\hat\beta)^{\mathsf{T}}\sum_{jii'} N_{\bullet j}^{-1} Z_{ij}Z_{i'j}(x_{ij}-x_{i'j})(a_i-a_{i'})\\
&+ (\beta-\hat\beta)^{\mathsf{T}}\sum_{jii'} N_{\bullet j}^{-1} Z_{ij}Z_{i'j}(x_{ij}-x_{i'j})(e_{ij}-e_{i'j}), \quad \textit{and}\\
U_e(\hat\beta) = U_e(\beta) &+ \frac{(\beta-\hat\beta)^{\mathsf{T}}}{2}\Bigl(\sum_{iji'j'} Z_{ij}Z_{i'j'}(x_{ij}-x_{i'j'})(x_{ij}-x_{i'j'})^{\mathsf{T}}\Bigr)(\beta-\hat\beta)\\
&+ (\beta-\hat\beta)^{\mathsf{T}}\sum_{iji'j'} Z_{ij}Z_{i'j'}(x_{ij}-x_{i'j'})(a_i-a_{i'})
\end{aligned}
$$

$$+ (\beta - \hat{\beta})^\mathsf{T} \sum_{iji'j'} Z_{ij} Z_{i'j'} (x_{ij} - x_{i'j'})(b_j - b_{j'})$$
$$+ (\beta - \hat{\beta})^\mathsf{T} \sum_{iji'j'} Z_{ij} Z_{i'j'} (x_{ij} - x_{i'j'})(e_{ij} - e_{i'j'}),$$

where for $\eta_{ij} = a_i + b_j + e_{ij}$,

$$U_a = \sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'} (\eta_{ij} - \eta_{ij'})^2,$$
$$U_b = \sum_{jii'} N_{\bullet j}^{-1} Z_{ij} Z_{i'j} (\eta_{ij} - \eta_{i'j})^2, \quad \textit{and}$$
$$U_e = \sum_{iji'j'} Z_{ij} Z_{i'j'} (\eta_{ij} - \eta_{i'j'})^2.$$

Proof. Straightforward algebra. □

Note that the η_{ij} exactly follow a two-factor crossed random effects model. Thus, Lemma 1 shows that we can leverage results about U_a, U_b, and U_e from [2] to analyze $U_a(\hat{\beta})$, $U_b(\hat{\beta})$, and $U_e(\hat{\beta})$.

S3.1 Proof of Theorem 4

Let the data be ordered by row and write $Y = X\beta + \eta$, where η has mean zero and variance $\sigma_A^2 A_R + \sigma_B^2 B_R + \sigma_E^2 I_N$. Then $\hat{\beta}_{\mathrm{OLS}} = \beta + (X^\mathsf{T} X)^{-1} X^\mathsf{T} \eta$. Clearly $\mathbb{E}((X^\mathsf{T} X)^{-1} X^\mathsf{T} \eta) = 0$. Now let $w \in \mathbb{R}^d$ be any unit vector. Then using matrices A_R and B_R from Section 3.1,

$$\mathrm{Var}(w^\mathsf{T} (X^\mathsf{T} X)^{-1} X^\mathsf{T} \eta)$$
$$= w^\mathsf{T} (X^\mathsf{T} X)^{-1} X^\mathsf{T} (\sigma_A^2 A_R + \sigma_B^2 B_R + \sigma_E^2 I_N) X (X^\mathsf{T} X)^{-1} w$$

$$
\begin{aligned}
&\leqslant (\sigma_E^2 + \sigma_A^2 \max_i N_{i\bullet} + \sigma_B^2 \max_j N_{\bullet j}) w^\mathsf{T} (X^\mathsf{T} X)^{-1} X^\mathsf{T} X (X^\mathsf{T} X)^{-1} w \\
&= \frac{1}{N} (\sigma_E^2 + \sigma_A^2 \max_i N_{i\bullet} + \sigma_B^2 \max_j N_{\bullet j}) w^\mathsf{T} \Bigl(\frac{1}{N} X^\mathsf{T} X \Bigr)^{-1} w \\
&\leqslant \Bigl(\frac{\sigma_E^2}{N} + \epsilon_R \sigma_A^2 + \epsilon_C \sigma_B^2 \Bigr) \Big/ \mathcal{I}\Bigl(\frac{X^\mathsf{T} X}{N} \Bigr) \\
&\to 0.
\end{aligned}
$$

The first inequality follows from the facts that the maximum eigenvalue of $\sigma_E^2 I_N$ is σ_E^2, the maximum eigenvalue of $\sigma_A^2 A_R$ is $\sigma_A^2 \max_i N_{i\bullet}$, and the maximum eigenvalue of $\sigma_B^2 B_R$ is $\sigma_B^2 \max_j N_{\bullet j}$. The conclusion now follows because $\max(1/N, \epsilon_R, \epsilon_C) \to 0$.

S3.2 Proof of Theorem 5

In light of Theorem 1 it suffices to show that $(U_a(\hat\beta) - U_a(\beta))/(N-R)$, $(U_b(\hat\beta) - U_b(\beta))/(N-C)$, and $(U_e(\hat\beta) - U_e(\beta))/N^2$ all converge to zero in probability. Because $\max(R,C)/N < \theta \in (0,1)$ we can replace denominators $N-R$ and $N-C$ by N. Using the expansion in Lemma 1,

$$
\begin{aligned}
&\frac{U_a(\hat\beta) - U_a(\beta)}{N} \\
&= \underbrace{\frac{1}{2} (\beta - \hat\beta)^\mathsf{T} \Bigl(\frac{1}{N} \sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'} (x_{ij} - x_{ij'})(x_{ij} - x_{ij'})^\mathsf{T} \Bigr) (\beta - \hat\beta)}_{\text{A1}} \\
&\quad + \underbrace{(\beta - \hat\beta)^\mathsf{T} \frac{1}{N} \sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'} (x_{ij} - x_{ij'})(b_j - b_{j'})}_{\text{A2}}
\end{aligned}
$$

$$\underbrace{+ (\beta - \hat\beta)^\mathsf{T} \frac{1}{N} \sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'} (x_{ij} - x_{ij'})(e_{ij} - e_{ij'})}_{\text{A3}}\,.$$

We consider Terms A1 through A3 in turn.

A1: The middle factor in Term A1 is no larger than $(4M_N/N)\sum_{ijj'} N_{i\bullet}^{-1} Z_{ij} Z_{ij'} = 4M_N = O(1)$. Therefore term A1 is $O(\|\beta - \hat\beta\|^2)$.

A2: The coefficient of $\hat\beta - \beta$ has mean zero and second moment

$$\begin{aligned}
&\mathbb{E}\Big(\frac{1}{N^2}\sum_{ijj'}\sum_{rss'} N_{i\bullet}^{-1} N_{r\bullet}^{-1} Z_{ij} Z_{ij'} Z_{rs} Z_{rs'} (x_{ij} - x_{ij'})(x_{rs} - x_{rs'})^\mathsf{T} \\
&\qquad\qquad (b_j - b_{j'})(b_s - b_{s'})\Big) \\
&= \frac{1}{N^2}\sum_{ijj'}\sum_{rss'} N_{i\bullet}^{-1} N_{r\bullet}^{-1} Z_{ij} Z_{ij'} Z_{rs} Z_{rs'} (x_{ij} - x_{ij'})(x_{rs} - x_{rs'})^\mathsf{T} \\
&\qquad \mathbb{E}((b_j - b_{j'})(b_s - b_{s'})) \\
&= \frac{\sigma_B^2}{N^2}\sum_{ijj'}\sum_{rss'} N_{i\bullet}^{-1} N_{r\bullet}^{-1} Z_{ij} Z_{ij'} Z_{rs} Z_{rs'} (x_{ij} - x_{ij'})(x_{rs} - x_{rs'})^\mathsf{T} \\
&\qquad (1_{j=s} - 1_{j=s'} - 1_{j'=s} + 1_{j'=s'}) \\
&= \frac{4\sigma_B^2}{N^2}\sum_{ijj'}\sum_{rs'} N_{i\bullet}^{-1} N_{r\bullet}^{-1} Z_{ij} Z_{ij'} Z_{rj} Z_{rs'} (x_{ij} - x_{ij'})(x_{rj} - x_{rs'})^\mathsf{T}.
\end{aligned}$$

No component in this matrix is larger than

$$\begin{aligned}
&\frac{16 M_N \sigma_B^2}{N^2}\sum_{ijj'}\sum_{rs'} N_{i\bullet}^{-1} N_{r\bullet}^{-1} Z_{ij} Z_{ij'} Z_{rj} Z_{rs'} \\
&= \frac{16 M_N \sigma_B^2}{N^2}\sum_{ij}\sum_{r} Z_{ij} Z_{rj} = \frac{16 M_N \sigma_B^2}{N^2}\sum_{j} N_{\bullet j}^2 = O(\epsilon),
\end{aligned}$$

and so Term A2 is $O(\|\hat\beta-\beta\|\epsilon)$.

A3: As in A2 we find that the coefficient of $\hat\beta-\beta$ has mean zero and second moment

$$\frac{4\sigma_E^2}{N^2}\sum_{ijj'}\sum_{s'}N_{i\bullet}^{-2}Z_{ij}Z_{ij'}Z_{is'}(x_{ij}-x_{ij'})(x_{ij}-x_{is'})^{\mathsf{T}}$$

in which no component is larger than

$$\frac{16\sigma_E^2M_N}{N^2}\sum_{ijj'}\sum_{s'}N_{i\bullet}^{-2}Z_{ij}Z_{ij'}Z_{is'}=\frac{16\sigma_E^2M_N}{N^2}\sum_{ij}Z_{ij}=\frac{16\sigma_E^2M_N}{N}.$$

Therefore Term A3 is $O(\|\hat\beta-\beta\|/N)$.

Combining these results $(U_a(\hat\beta)-U_a(\beta))/N=O(\|\hat\beta-\beta\|(\epsilon+\|\hat\beta-\beta\|))$. The same argument applies to $(U_b(\hat\beta)-U_b(\beta))/N$. Now we turn to $(U_e(\hat\beta)-U_e(\beta))/N^2$. Using the expansion in Lemma 1,

$$\begin{aligned}\frac{U_e(\hat\beta)-U_e(\beta)}{N^2}&=\underbrace{\frac{(\beta-\hat\beta)^{\mathsf{T}}}{2}\Bigl(\frac{1}{N^2}\sum_{iji'j'}Z_{ij}Z_{i'j'}(x_{ij}-x_{ij'})(x_{ij}-x_{ij'})^{\mathsf{T}}\Bigr)(\beta-\hat\beta)}_{\text{E1}}\\&\quad+\underbrace{(\beta-\hat\beta)^{\mathsf{T}}\frac{1}{N^2}\sum_{iji'j'}Z_{ij}Z_{i'j'}(x_{ij}-x_{i'j'})(a_i-a_{i'})}_{\text{E2}}\\&\quad+\underbrace{(\beta-\hat\beta)^{\mathsf{T}}\frac{1}{N^2}\sum_{iji'j'}Z_{ij}Z_{i'j'}(x_{ij}-x_{i'j'})(b_j-b_{j'})}_{\text{E3}}\\&\quad+\underbrace{(\beta-\hat\beta)^{\mathsf{T}}\frac{1}{N^2}\sum_{iji'j'}Z_{ij}Z_{i'j'}(x_{ij}-x_{i'j'})(e_{ij}-e_{i'j'})\,.}_{\text{E4}}\end{aligned}$$

E1: By arguments like the one for A1, we find that E1 is also $O(\|\hat{\beta}-\beta\|^2)$.

E2: Similarly to A2, the coefficient of $\hat{\beta}-\beta$ has mean zero and second moment

$$\mathbb{E}\Big(\frac{1}{N^4}\sum_{iji'j'}\sum_{rsr's'} Z_{ij}Z_{i'j'}Z_{rs}Z_{r's'}(x_{ij}-x_{i'j'})(x_{rs}-x_{r's'})^\mathsf{T}(a_i-a_{i'})(a_r-a_{r'})\Big)$$
$$=\frac{4\sigma_A^2}{N^4}\sum_{iji'j'}\sum_{sr's'} Z_{ij}Z_{i'j'}Z_{is}Z_{r's'}(x_{ij}-x_{i'j'})(x_{is}-x_{r's'})^\mathsf{T}$$

with components no larger than

$$\frac{16M_N\sigma_A^2}{N^4}\sum_{iji'j'}\sum_{sr's'} Z_{ij}Z_{i'j'}Z_{is}Z_{r's'}$$
$$=\frac{16M_N\sigma_A^2}{N^2}\sum_{ij}\sum_{s} Z_{ij}Z_{is}=\frac{16M_N\sigma_A^2}{N^2}\sum_{i} N_{i\bullet}^2.$$

Therefore Term E2 is $O_p(\|\hat{\beta}-\beta\|\epsilon)$.

E3: Term E3 is also $O_p(\|\hat{\beta}-\beta\|\epsilon)$. by the argument used for Term E3.

E4: Following arguments similar to the preceding ones, the coefficient of $\hat{\beta}-\beta$ has mean zero and second moment

$$\frac{4\sigma_E^2}{N^4}\sum_{iji'j'}\sum_{r's'} Z_{ij}Z_{i'j'}Z_{r's'}(x_{ij}-x_{i'j'})(x_{ij}-x_{r's'})^\mathsf{T}=O\Big(\frac{16\sigma_E^2M_N}{N}\Big)$$

and so Term E4 is $O(\|\hat{\beta}-\beta\|/N)$. Combining these results we have consistency for the variance components. The error in replacing β by $\hat{\beta}$ changes the variance component estimates by $O(\|\hat{\beta}-\beta\|(\|\hat{\beta}-\beta\|+\epsilon))$.

S3.3 Proof of Theorem 6

Suppose that the data are ordered by rows. Then we may write $Y = X\beta + \eta$, where η has mean zero and variance $V_R = \sigma_E^2 I_N + \sigma_A^2 A_R + \sigma_B^2 B_R$. Now

$$\hat\beta_{\mathrm{RLS}} = \beta + (X^\mathsf{T}\hat V_A^{-1}X)^{-1}X^\mathsf{T}\hat V_A^{-1}\eta$$

where $\hat V_A = \hat\sigma_A^2 A_R + \hat\sigma_E^2 I_N$. The matrix X is not random and both $\hat\sigma_A^2 \xrightarrow{p} \sigma_A^2$ and $\hat\sigma_E^2 \xrightarrow{p} \sigma_E^2$ so it suffices to show that $\varepsilon \equiv (X^\mathsf{T}V_A^{-1}X)^{-1}X^\mathsf{T}V_A^{-1}\eta \xrightarrow{p} 0$. Write $\eta = a + b + e$ where these are the random effects in the row order. We can easily handle the effect of $a + e$, via

$$\begin{aligned}\mathrm{Var}((X^\mathsf{T}V_A^{-1}X)^{-1}X^\mathsf{T}V_A^{-1}(a+e)) &= (X^\mathsf{T}V_A^{-1}X)^{-1}X^\mathsf{T}V_A^{-1}X(X^\mathsf{T}V_A^{-1}X)^{-1}\\ &= (X^\mathsf{T}V_A^{-1}X)^{-1}.\end{aligned}$$

The largest eigenvalue of V_A is $O(N_{i\bullet})$ and so this quantity is $O(\epsilon_R) \to 0$.

We will need a sharper analysis of $(X^\mathsf{T}V_A^{-1}X)^{-1}$ to control the contribution of the column random effects b to the row-weighted GLS estimate $\hat\beta_{\mathrm{RLS}}$. Furthermore their contribution to the intercept term in β motivates centering the x_{ij}. For a nonrandom invertible matrix $K \in \mathbb{R}^{p\times p}$, we may replace X by $X^* = XK$ and β by $\beta^* = K^{-1}\beta$. Now $\hat\beta_{\mathrm{RLS}} = K\hat\beta^*_{\mathrm{RLS}}$ and so $\mathrm{Var}(\hat\beta_{\mathrm{RLS}}) = K\mathrm{Var}(\hat\beta^*_{\mathrm{RLS}})K^\mathsf{T}$. Our matrix K will be uniformly bounded as $N \to \infty$ and independent of η. Then $\mathrm{Var}(\hat\beta^*_{\mathrm{RLS}}) \to 0$ implies $\mathrm{Var}(\hat\beta_{\mathrm{RLS}}) \to 0$.

The matrix we choose is

$$K = \begin{pmatrix} 1 & -k_2 & \cdots & -k_p \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

with values k_t for $t = 2, \dots, p$ given below. We have $x_{ij}^* = (1, x_{ij,2} - k_2, x_{ij,3} - k_3, \dots, x_{ij,p} - k_p)$.

We begin by noting that in the row ordering,

$$V_A^{-1} = \frac{1}{\sigma_E^2} \mathrm{diag}\Bigl(I_{N_{i\bullet}} - N_{i\bullet}^{-1} \gamma_i 1_{N_{i\bullet}} 1_{N_{i\bullet}}^\mathsf{T}\Bigr)$$

where there are R diagonal blocks of size $N_{i\bullet} \times N_{i\bullet}$ and $\gamma_i = N_{i\bullet}\sigma_A^2/(\sigma_E^2 + N_{i\bullet}\sigma_A^2)$. Then

$$\sigma_E^2 X^\mathsf{T} V_A^{-1} X = X^\mathsf{T}\bigl(X - \mathrm{col}\bigl(\gamma_i 1_{N_{i\bullet}} \bar{x}_{i\bullet}^\mathsf{T}\bigr)\bigr)$$

where $\mathrm{col}(\cdot) \in \mathbb{R}^{N \times p}$ is a column of R blocks of sizes $N_{i\bullet} \times p$. Continuing

$$\begin{aligned} \sigma_E^2 X^\mathsf{T} V_A^{-1} X &= \sum_i \sum_j Z_{ij}(x_{ij} x_{ij}^\mathsf{T} - x_{ij} \bar{x}_{i\bullet}^\mathsf{T} \gamma_i) \\ &= X^\mathsf{T} X - \sum_i \gamma_i \sum_j Z_{ij} x_{ij} \bar{x}_{i\bullet}^\mathsf{T} \\ &= X^\mathsf{T} X - \sum_i N_{i\bullet} \gamma_i \bar{x}_{i\bullet} \bar{x}_{i\bullet}^\mathsf{T} \\ &= \sum_{ij} Z_{ij}(x_{ij} - \bar{x}_{i\bullet})(x_{ij} - \bar{x}_{i\bullet})^\mathsf{T} + \sum_i N_{i\bullet}(1 - \gamma_i) \bar{x}_{i\bullet} \bar{x}_{i\bullet}^\mathsf{T}. \end{aligned} \tag{S3.9}$$

The lower right $(p-1)\times(p-1)$ submatrix of the first term in (S3.9) grows proportionally to N. We will see that the upper left 1×1 submatrix of the second term grows at least as fast as R. We choose our matrix K to zero out all of the top row and leftmost column of $X^{*\mathsf{T}}V_A^{-1}X^*$ except the upper left entry. To this end, define

$$k_t = \frac{\sum_i N_{i\bullet}(1-\gamma_i)\bar{x}_{i\bullet,t}}{\sum_i N_{i\bullet}(1-\gamma_i)} = \frac{\sum_i x_{i\bullet,t}\, N_{i\bullet}\sigma_E^2/(N_{i\bullet}\sigma_A^2+\sigma_E^2)}{\sum_i N_{i\bullet}\sigma_E^2/(N_{i\bullet}\sigma_A^2+\sigma_E^2)}, \quad t=2,\dots,p.$$

Now from (S3.9),

$$\sigma_E^2 X^{*\mathsf{T}}V_A^{-1}X^* = \begin{pmatrix} \sum_i N_{i\bullet}(1-\gamma_i) & 0_{p-1}^{\mathsf{T}} \\ 0_{p-1} & V \end{pmatrix}$$

where V is the lower right $(p-1)\times(p-1)$ submatrix of $\sum_{ij} Z_{ij}(x_{ij}-\bar{x}_{i\bullet})(x_{ij}-\bar{x}_{i\bullet})^{\mathsf{T}}$ plus a positive semidefinite matrix. Therefore

$$\bigl(X^{*\mathsf{T}}V_A^{-1}X^*\bigr)^{-1} = \sigma_E^2 \begin{pmatrix} 1/\sum_i N_{i\bullet}(1-\gamma_i) & 0_{p-1}^{\mathsf{T}} \\ 0_{p-1} & V^{-1} \end{pmatrix}.$$

Continuing the derivation,

$$\begin{aligned} &\mathrm{Var}((X^{*\mathsf{T}}V_A^{-1}X^*)^{-1}X^{*\mathsf{T}}V_A^{-1}b) \\ &= (X^{*\mathsf{T}}V_A^{-1}X^*)^{-1}(X^{*\mathsf{T}}V_A^{-1}B_RV_A^{-1}X^*)(X^{*\mathsf{T}}V_A^{-1}X^*)^{-1}\sigma_B^2. \end{aligned}$$

The eigenvalues of V_A^{-1} are all smaller than 1, so in the ordering of positive semidefinite matrices,

$$X^{*\mathsf{T}}V_A^{-1}B_RV_A^{-1}X^* \leqslant X^{*\mathsf{T}}B_RX^* = \sum_j N_{\bullet j}^2\bar{x}_{\bullet j}^*\bar{x}_{\bullet j}^{*\mathsf{T}}.$$

Now for a unit vector $w \in \mathbb{R}^p$ with $w_1 = 0$ we have $\|(X^{*\mathsf{T}} V_A^{-1} X^*)^{-1} w\| \leqslant cN^{-1}\sigma_E^2$ because the sample covariance of non-intercept x's grows (at least) proportionally to N. The x_{ij} are bounded and so therefore the $\bar{x}^*_{\bullet j}$ are also bounded. So now

$$\mathrm{Var}(w^{\mathsf{T}}(X^{*\mathsf{T}} V_A^{-1} X^*)^{-1} X^{*\mathsf{T}} V_A^{-1} b) = O(N^{-2} \sum_j N_{\bullet j}^2) \to 0.$$

Next we consider $w = (1, 0, \dots, 0)$. For this w,

$$(X^{*\mathsf{T}} V_A^{-1} X^*)^{-1} w = \frac{\sigma_E^2}{\sum_i N_{i\bullet}(1-\gamma_i)} = \frac{\sigma_E^2}{\sum_i N_{i\bullet}\sigma_E^2/(N_{i\bullet}\sigma_A^2 + \sigma_E^2)} \leqslant \frac{1}{R\sigma_A^2}.$$

The matrix B_R has C blocks of the form $1_{N_{\bullet j}} 1_{N_{\bullet j}}^{\mathsf{T}}$ permuted into the row ordering. We may write $B_R = Z_b Z_b^{\mathsf{T}}$ where $Z_b \in \{0,1\}^{N \times C}$. The row of Z_b corresponding to observation ij has only one 1 in it, at position j. Now

$$V_A^{-1} X^* = \begin{pmatrix} (1-\gamma_1) 1_{N_{1\bullet}} & 0 & 0 & \cdots & 0 \\ (1-\gamma_2) 1_{N_{2\bullet}} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-\gamma_R) 1_{N_{R\bullet}} & 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{N \times p}$$

and the j'th row of $Z_b V_A^{-1} X^* \in \mathbb{R}^{C \times p}$ is $(\sum_i Z_{ij}(1-\gamma_i), 0, \dots, 0) \in \mathbb{R}^p$. Then the only nonzero element of $(X^{*\mathsf{T}} V_A^{-1} B_R V_A^{-1} X^*)$ is the upper left one and it equals $\sum_{ijr} Z_{ij} Z_{rj} (1-\gamma_i)(1-\gamma_r)$. Therefore, for $w = (1, 0, \dots, 0)$,

$$\mathrm{Var}(w^{\mathsf{T}}(X^{*\mathsf{T}} V_A^{-1} X^*)^{-1} X^{*\mathsf{T}} V_A^{-1} b) \leqslant \frac{1}{R^2 \sigma_A^4} \sum_{ijr} Z_{ij} Z_{rj} (1-\gamma_i)(1-\gamma_r)$$

$$
\begin{aligned}
&= \frac{1}{R^2\sigma_A^4}\sum_{ir}(ZZ^\mathsf{T})_{ir}\frac{\sigma_E^2}{\sigma_E^2+N_{i\bullet}\sigma_A^2}\frac{\sigma_E^2}{\sigma_E^2+N_{r\bullet}\sigma_A^2}\\
&\leqslant \frac{\sigma_E^4}{R^2\sigma_A^8}\sum_{ir}(ZZ^\mathsf{T})_{ir}N_{i\bullet}^{-1}N_{r\bullet}^{-1}
\end{aligned}
$$

which vanishes by condition (5.14). A general unit vector w can be written as a linear combination of unit vectors with $w_1 = 0$ and $w_1 = 1$ and so $\mathrm{Var}(w^\mathsf{T}(X^{*\mathsf{T}}V_A^{-1}X^*)^{-1}X^{*\mathsf{T}}V_A^{-1}b) \to 0$. Because K is bounded $\mathrm{Var}(w^\mathsf{T}(X^\mathsf{T}V_A^{-1}X)^{-1}X^\mathsf{T}V_A^{-1}b) \to 0$ as well. This completes the proof.

S3.4 Proof of Theorem 7

We will use the following central limit theorem for a triangular array of weighted sums of IID random variables.

Theorem 2. *For integers i and n with $1 \leqslant i \leqslant n$, let $\epsilon_{n,i}$ be a triangular array of random variables that are IID within each row with mean μ_n and variance $\sigma_n^2 \in (0,\infty)$. Let $c_{n,i}$ be a triangular array of finite constants, not all zero within each row. Define*

$$
T_n = \frac{1}{B_n}\sum_{i=1}^n c_{ni}(\epsilon_{n,i}-\mu_n), \quad \textit{where} \quad B_n^2 = \sigma_n^2\sum_{i=1}^n c_{ni}^2.
$$

If $\max_{1\leqslant i\leqslant n} c_{ni}^2/\sum_{i=1}^n c_{ni}^2 \to 0$ as $n\to\infty$, then $T_n \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.

Proof. This is from Theorem 2.2 of [1]. □

Our use case is for $\mu_n = 0$ and σ_n constant in n. That case was also handled by [3, Theorem 1] who has a converse.

From Section S3.3, $\hat{\beta}_{\mathrm{RLS}} - \beta = (X^\mathsf{T} V_A^{-1} X)^{-1} X^\mathsf{T} V_A^{-1} \eta$, where $\eta_{ij} = a_i + b_j + e_{ij}$. We will make use of sums $\eta_{i\bullet} = \sum_j Z_{ij}\eta_{ij}$ and $X_{i\bullet} = \sum_j Z_{ij} x_{ij} \in \mathbb{R}^p$ as well as corresponding column sums. The matrix $(X^\mathsf{T} V_A^{-1} X)^{-1}$ is not random. We will establish a central limit theorem for $X^\mathsf{T} V_A^{-1}\eta$.

Consider $w^\mathsf{T} X^\mathsf{T} V_A^{-1}\eta$ for a unit vector $w \in \mathbb{R}^p$. By the Woodbury formula,

$$
\begin{aligned}
w^\mathsf{T} X^\mathsf{T} V_A^{-1}\eta &= \frac{w^\mathsf{T} X^\mathsf{T}\eta}{\sigma_E^2} - \frac{\sigma_A^2}{\sigma_E^2}\sum_i \frac{w^\mathsf{T} X_{i\bullet}\eta_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}} \\
&= \frac{1}{\sigma_E^2}\Bigl[\sum_i a_i w^\mathsf{T} X_{i\bullet} + \sum_j b_j w^\mathsf{T} X_{\bullet j} + \sum_{ij} Z_{ij} e_{ij} w^\mathsf{T} x_{ij}\Bigr] \\
&\quad - \frac{\sigma_A^2}{\sigma_E^2}\sum_i \frac{w^\mathsf{T} X_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}\Bigl(N_{i\bullet} a_i + \sum_j Z_{ij} b_j + \sum_j Z_{ij} e_{ij}\Bigr) \\
&= \underbrace{\sum_i a_i \frac{w^\mathsf{T} X_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}}_{\text{Term R1}} + \underbrace{\sum_j \frac{b_j}{\sigma_E^2}\Bigl(w^\mathsf{T} X_{\bullet j} - \sigma_A^2 \sum_i Z_{ij} \frac{w^\mathsf{T} X_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}\Bigr)}_{\text{Term R2}} \\
&\quad + \underbrace{\sum_{ij} Z_{ij}\frac{e_{ij}}{\sigma_E^2}\Bigl(w^\mathsf{T} x_{ij} - \sigma_A^2 \frac{w^\mathsf{T} X_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}\Bigr)}_{\text{Term R3}}.
\end{aligned}
$$

Terms R1, R2 and R3 are independent. We will show CLTs for each of them individually.

R1: We use Theorem 2 with random variables a_i and weights $c_i = w^\mathsf{T} X_{i\bullet}/(\sigma_E^2 + \sigma_A^2 N_{i\bullet})$. Now $\max_i c_i^2 \leqslant M_N^2$ and

$$
\sum_i c_i^2 = \sum_i \Bigl(\frac{w^\mathsf{T} X_{i\bullet}}{\sigma_E^2 + \sigma_A^2 N_{i\bullet}}\Bigr)^2 \geqslant \sum_i \Bigl(\frac{w^\mathsf{T} \bar{x}_{i\bullet}}{\sigma_E^2 + \sigma_A^2}\Bigr)^2
$$

$$\geqslant (\sigma_E^2+\sigma_A^2)^{-2}\,\mathcal{I}\biggl(\sum_i \bar{x}_{i\bullet}\bar{x}_{i\bullet}^{\mathsf{T}}\biggr)\to\infty.$$

Therefore Term R1 is asymptotically normally distributed.

R2: This term is a weighted sum of independent random variables b_j/σ_E^2 with weights $c_j = w^{\mathsf{T}}\sum_i Z_{ij}(x_{ij}-\gamma_i\bar{x}_{i\bullet})$, where $\gamma_i = \sigma_A^2/(\sigma_A^2+\sigma_E^2/N_{i\bullet})$. Therefore $c_j = N_{\bullet j}w^{\mathsf{T}}(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})$ for the second order averages $\tilde{x}_{\bullet j}$ given by (5.15).

As in the proof of Theorem 7 from Section S3.4 we employ a bounded invertible centering matrix $K = \bigl(\begin{smallmatrix}1 & -k\\ 0 & I_{p-1}\end{smallmatrix}\bigr)$, not necessarily the same one as there. We will show that $K\sum_j N_{\bullet j}b_j(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})$ is asymptotically Gaussian and then so is $\sum_j N_{\bullet j}b_j(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})$. Let $c_j^* = w^{\mathsf{T}}K\sum_j N_{\bullet j}(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})$. Then

$$\sum_j c_j^{*2} = w^{\mathsf{T}}\sum_j N_{\bullet j}^2 K(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j})^{\mathsf{T}}K^{\mathsf{T}}w.$$

For $2\leqslant t\leqslant p$ let

$$k_t = \sum_j N_{\bullet j}^2(\bar{x}_{\bullet j,t}-\tilde{x}_{\bullet j,t})\Bigm/\sum_j N_{\bullet j}^2$$

and define $k^* = (0,k_2,\dots,k_p)^{\mathsf{T}}$. Then $\sum_j N_{\bullet j}^2 K(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j}-k^*)(\bar{x}_{\bullet j}-\tilde{x}_{\bullet j}-k^*)^{\mathsf{T}}K^{\mathsf{T}}$ is block diagonal with an upper 1×1 block and a lower $(p-1)\times(p-1)$ block.

First suppose that $w=(w_1,w_2,\dots,w_p)$ is a unit vector with $|w_1|\neq 1$.

Then,

$$\sum_j c_j^{*2} \geqslant \|w\|_{-1}^2 \mathcal{I}_0\Big(\sum_j N_{\bullet j}^2 (\bar{x}_{\bullet j} - \tilde{x}_{\bullet j} - k^*)(\bar{x}_{\bullet j} - \tilde{x}_{\bullet j} - k^*)^{\mathsf{T}}\Big)$$

which diverges faster than $\max_j N_{\bullet j}^2$ by hypothesis. It remains to consider $w^{\mathsf{T}} = (\pm 1, 0, \ldots, 0)$. For this vector $c_j^* = c_j = \sum_i Z_{ij}(1 - \gamma_i) = \sum_i Z_{ij}\sigma_E^2/(\sigma_E^2 + N_{i\bullet}\sigma_A^2)$ and $\max_j c_j^2 / \sum_j c_j^2 \to 0$ by hypothesis. Therefore Term R2 is asymptotically normally distributed.

R3: This term is a weighted sum of IID random variables e_{ij}/σ_E^2 with weights $c_{ij} = Z_{ij} w^{\mathsf{T}}(x_{ij} - \gamma_i \bar{x}_{i\bullet})$. As in paragraph R2, we employ a bounded invertible centering matrix. Then for a unit vector $w \neq (\pm 1, 0, \ldots, 0)^{\mathsf{T}}$

$$\begin{aligned}\sum_{ij} {c_{ij}^*}^2 &\geqslant \|w\|_{-1}^2 \mathcal{I}_0\Big(\sum_{ij} Z_{ij}(x_{ij} - \gamma_i \bar{x}_{i\bullet})(x_{ij} - \gamma_i \bar{x}_{i\bullet})^{\mathsf{T}}\Big)\\ &\geqslant \|w\|_{-1}^2 \mathcal{I}_0\Big(\sum_{ij} Z_{ij}(x_{ij} - \bar{x}_{i\bullet})(x_{ij} - \bar{x}_{i\bullet})^{\mathsf{T}}\Big)\end{aligned}$$

which, by hypothesis, diverges to infinity, while $\max_{ij} {c_{ij}^*}^2 = O(1)$. The case $w = (\pm 1, 0, \ldots, 0)^{\mathsf{T}}$ is handled by one of the assumptions in the theorem.

All three terms have asymptotic normal distributions with mean zero, and they are independent. Therefore, $(X^{\mathsf{T}} V_A^{-1} X)^{-1} X^{\mathsf{T}} V_A^{-1} \eta$ is asymptotically Gaussian with mean zero and variance

$(X^{\mathsf{T}} V_A^{-1} X)^{-1} X^{\mathsf{T}} V_A^{-1} V_R V_A^{-1} X (X^{\mathsf{T}} V_A^{-1} X)^{-1}$.

Table 1: Meanings of the column headings in the regression output.

var	Coefficient name
bhatols	$\hat{\beta}_{\mathrm{OLS}}$, the OLS regression coefficient
selhslhs	$\mathrm{SE}_{\mathrm{OLS}}(\hat{\beta}_{\mathrm{OLS}})$, standard error from OLS formula
selhs	$\mathrm{SE}_{\mathrm{Moments}}(\hat{\beta}_{\mathrm{OLS}})$, standard error of OLS coefficients using moments
bhat	$\hat{\beta}_{\mathrm{Moments}}$, the method of moments coefficient
se	$\mathrm{SE}_{\mathrm{Moments}}(\hat{\beta}_{\mathrm{Moments}})$, the method of moments standard error

S4 Regression coefficients from Stitch Fix example

This section has the regression output for the Stitch Fix regression example for all regression variables. The column headings are explained in Table 1. Here Cedgy refers to the client being edgy, Iedgy describes the item being and Bedgy indicates that both are edgy, that is Bedgy=Cedgy×Iedgy. Boho is treated similarly. Here is the full table, verbatim.

var	bhatols	selhslhs	selhs	bhat	se
Intercept	4.635000	0.005397	0.058080	5.110000	0.012500
Match	5.048000	0.011740	0.146400	3.529000	0.021530
Cedgy	0.001020	0.002443	0.004593	0.001860	0.003831
Iedgy	-0.335800	0.004253	0.037300	-0.332800	0.015420
Bedgy	0.392500	0.006229	0.013520	0.386400	0.006432
Cboho	0.138600	0.002264	0.004354	0.133400	0.003622

Iboho	-0.549900	0.005981	0.030490	-0.626100	0.016610
Bboho	0.382200	0.007566	0.010570	0.383700	0.007697
Acrylic	-0.064820	0.003778	0.038040	-0.016270	0.021490
Angora	-0.012620	0.007848	0.096310	0.072710	0.058370
Bamboo	-0.045930	0.062150	0.243700	0.054200	0.171600
Cashmere	-0.195500	0.024840	0.159300	0.013540	0.117600
Cotton	0.175200	0.003172	0.047660	0.097430	0.018110
Cupro	0.597900	0.301600	0.485700	0.560300	0.485200
FauxFur	0.275900	0.020080	0.086310	0.364900	0.075240
Fur	-0.202100	0.031210	0.156000	-0.034780	0.133100
Leather	0.267700	0.024820	0.086710	0.279800	0.073350
Linen	-0.384400	0.056320	0.272900	0.006269	0.166000
Modal	0.002587	0.009775	0.205200	0.141700	0.064980
Nylon	0.033490	0.015520	0.100000	0.118600	0.064360
PatentLeather	-0.235900	0.180000	0.423500	-0.247300	0.422200
Pleather	0.416300	0.008916	0.099050	0.334400	0.050230
PU	0.416000	0.008225	0.090190	0.495100	0.041960
PVC	0.657400	0.065450	0.389800	0.871300	0.388300
Rayon	-0.011090	0.002951	0.046020	0.010290	0.014930
Silk	-0.142200	0.013170	0.100400	-0.165600	0.054710

Spandex	0.391600	0.017290	0.154900	0.363100	0.128400
Tencel	0.496600	0.009313	0.193500	0.154800	0.067180
Viscose	0.040660	0.006953	0.096200	-0.013890	0.035270
Wool	-0.060210	0.006611	0.081410	-0.006051	0.037370

References

[1] Subhash C. Bagui, Dulal K. Bhaumik, and K. L. Mehra. A few counter examples useful in teaching central limit theorems. *The American Statistician*, 67(1):49–56, 2013.

[2] Katelyn Gao and Art B. Owen. Efficient moment calculations for variance components in large unbalanced crossed random effects models. *Electronic Journal of Statistics*, 11(1):1235–1296, 2017.

[3] Péter Kevei. A note on asymptotics of linear combinations of IID random variables. *Periodica Mathematica Hungarica*, 60(1):25–36, 2010.