Leveraging Large Language Models To Improve Precision in Randomized Controlled Trials

Jaylin Lowe[1], Adam Sales[2], and Johann Gagnon-Bartsch[1]

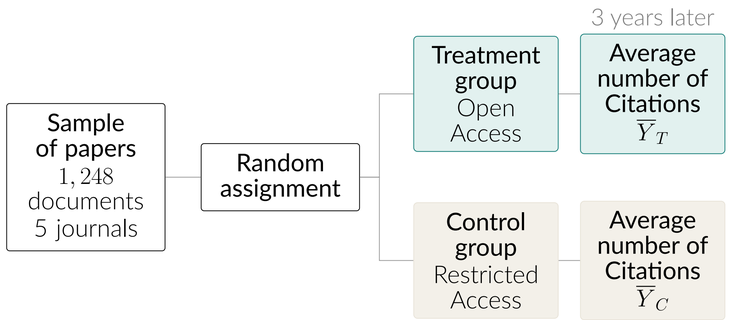
[1]University of Michigan [2]Worcester Polytechnic Institute

Motivation

Randomized controlled trials (RCTs) can be small and costly, yielding imprecise estimates. Covariate adjustment accounts for chance imbalances between groups, potentially improving precision. Observational data can provide predicted outcomes as an additional covariate.

What if we used a large language model (LLM) in place of an observational dataset?

Case Study: Open Access Paper Citations



Setup

Under the potential outcomes framework [2, 3] we have:

- N observations, indexed from $i = 1, \ldots N$
- Z, vector of randomized treatment assignments where $Z_i = 1$ denotes treatment and $Z_i = 0$ denotes control.
- p, the probability of being assigned to treatment
- y_i^t and y_i^c, the potential outcomes under treatment and control

External Data For Covariate Adjustment

We use the method outlined in [1] adapted to use LLM predictions:

(1) Obtain a vector of predictions $\hat{y}_i^{LLM}$ from an LLM.

(2) Augment the RCT covariates, $\mathbf{x_i}$, with the LLM-generated covariate: $\tilde{\mathbf{x}}_\mathbf{i} = [x_i, \hat{y}_i^{LLM}]$.

(3) Obtain imputations $\hat{y}^t(\tilde{\mathbf{x}}_\mathbf{i})$ and $\hat{y}^c(\tilde{\mathbf{x}}_\mathbf{i})$ for each RCT observation, independently of its treatment assignment.

(4) Calculate

$$\hat{\tau} = \frac{1}{N}\sum_{i=1}^{N} Z_i \cdot \frac{Y_i - \hat{m}_i}{p} - \frac{1}{N}\sum_{i=1}^{N}(1 - Z_i) \cdot \frac{Y_i - \hat{m}_i}{1-p} \quad (1)$$

where $\hat{m}_i = p\hat{y}^c(\tilde{\mathbf{x}}_\mathbf{i}) + (1-p)\hat{y}^t(\tilde{\mathbf{x}}_\mathbf{i})$.

$$\widehat{\mathrm{Var}}(\hat{\tau}) = \frac{1}{N}\left[\frac{p}{1-p}\hat{E}_c^2 + \frac{1-p}{p}\hat{E}_t^2 + 2\sqrt{\hat{E}_c^2\hat{E}_t^2}\right] \quad (2)$$

where $\hat{E}_c^2 = \frac{1}{n_c}\sum_{i=1}^{N}(1 - Z_i) * [y_i^c - \hat{y}_{-i}^c(\tilde{\mathbf{x}}_i)]^2$ and $\hat{E}_t^2$ is defined analogously.

Our Method

Goal: Obtain an additional covariate from the LLM that is predictive of the RCT outcome.

(1) Pair observations. To improve precision or reduce computation, use a subset of pairs, stratified by covariates.

(2) Based on the abstracts, ask the LLM which observation in a pair is more likely to exhibit the target quality.

(3) Convert these responses into selection frequencies to be used as additional covariates.

(4) Calculate the variance of the ATE with the new covariates.

Results

- Observations are stratified by journal.
- We compare four models: one with just the base covariates, one with base covariates plus the LLM citation based score, one with the base covariates plus the 10 LLM based quality scores, and one with all of these covariates.
- Also applied to two other case studies with only typical, rectangular covariates with no substantial improvement.

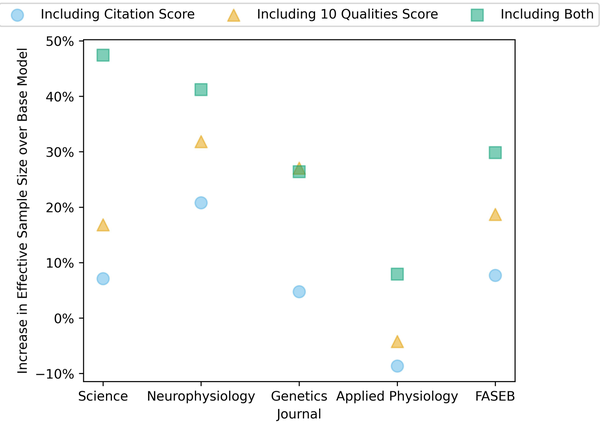


Figure 1. Equivalent Sample Size Improvement by Model Type and Journal

References

[1] J. A. Gagnon-Bartsch, A. C. Sales, E. Wu, A. F. Botelho, J. A. Erickson, L. W. Miratrix, and N. T. Heffernan. Precise unbiased estimation in randomized experiments using auxiliary observational data. *Journal of Causal Inference*, 11(1):20220011, Jan. 2023.

[2] J. Neyman. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1923.

[3] D. B. Rubin. Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

Poster created using template by David Gamba, University of Michigan.