THESIS PROPOSAL: Power Analysis in Conjoint Experiments

Alberto Stefanelli[[1]](#footnote-20).

**PROPOSED SUPERVISOR: Bart Meuleman**

# Problem Statement

In the social sciences, conjoint experiments have become a prominent methodological tool to measure causal effects of many attributes, modelling complex and multidimensional preferences. These include various phenomena, such as preferences for political candidates, immigrant admissions, or public policies. The fully randomized design introduced by Hainmueller, Hopkins, and Yamamoto ([2014](#ref-hainmueller_causal_2014)) was a marked change and a driving force behind the recent rise in conjoint experiments in political science, sociology, and economics. The quantity of interest—the average marginal component effect (AMCE) is available as a result of the randomized con- joint design that breaks any explicit or implicit confounding between features of the experiment.

The raising popularity of conjoint experiments is due to the possibility of manipulating multiple components (i.e., attributes) of interest and estimate the effects of an attribute at several levels of the other attributes. This design—usually called factorial design—have a long long tradition in the social sciences, but testing for numerous attributes at once quickly becomes infeasible due to the necessity of a large number of observations. Despite the rising popularity of conjoint studies, researchers often disregard this issue with the consequence that *a priori* power analysis is often ignored when designing conjoint experiments. For instance, scholars have argued that conjoint designs “free us from the power constraints that limit traditional factorial experiments” ([Kertzer, Renshon, and Yarhi-Milo 2019, 7](#ref-kertzer_how_2019)) and solve key problems in experimental research such as “the trade-off between statistical power and the desire to employ many experimental conditions” ([Knudsen and Johannesson 2018, 2](#ref-knudsen_beyond_2018)).

Power analysis consists of assessing the probability of successfully rejecting the null hypothesis when it is *false*. In other words, it can be used to estimate the required sample size to detect an effect of a given size with a certain degree of confidence and precision. Power analysis also reduces the chances of obtaining false-positives or exaggerated findings ([Gelman and Carlin 2014](#ref-gelman_carlin2014)). The difficulty in replicating some experimental studies or the emergence of contradictory results in the literature are caused—*inter alia*—by under-powered experimental designs ([Open Science Collaboration 2015](#ref-OSC2015); [Maxwell 2004](#ref-Maxwell2004)). The reason is that studies with low power are more likely to yield large and significant effects that are due to random fluctuations alone. Finally, *a priori* power analysis can save costs by cautioning against over-powered designs with a much higher number of respondents (and, thus, higher costs) than necessary for discovering the hypothesized effect.

Despite their growing popularity, calculating power for a conjoint experiment is not a trivial task. Compared to traditional survey experiments, conjoint applications present a relatively large number of design choices. In addition to sample and effect size considerations, experimentalists need to decide on the number of attributes (the features of the profiles, such as the gender of a hypothetical candidate), the number of levels of such attributes (the values of that attributes can take, such as male or female), and the number of tasks (the number of profile choices a respondent will undertake). This adds a layer of complexity to the analysis of statistical power with respect to the trade-off between the number of experimental conditions, the number of tasks that each respondent shall undertake, and the required sample size.

# Power Analysis for Conjoint Expeirments

Previous research on sample size requirements for conjoint experiments is slim and fails to provide accurate guidance for designing conjoint experiments ([Orme 1998](#ref-orme_sample_1998); [Louviere et al. 2000](#ref-louviere_stated_2000); [Rose and Bliemer 2013](#ref-rose_sample_2013)). The few works on the topic do not directly focus on the approach proposed by Hainmueller, Hopkins, and Yamamoto ([2014](#ref-hainmueller_causal_2014)) and typically disregard the number of attributes and levels included in the design (e.g., [Louviere et al. 2000](#ref-louviere_stated_2000); [Rose and Bliemer 2013](#ref-rose_sample_2013)). In addition, current literature provides no indications of the impact of design choices on the probability of an estimate being biased upward or downward (the *Type M error)* or being in the wrong direction (the *Type S error*). The only exception is Schuessler and Freitag ([2020](#ref-Schuessler2020)) who employ parametric techniques from traditional discrete choice experiment to derive minimal sample size for conjoint experiments. However, their approach cannot be easily expanded to accommodate more complex designs, do not distinguish between the number of respondents and the number of trials, and omits the cluster structure typically found in conjoint designs.

Therefore, this work aims at developing a flexible simulation-based framework based on the approach proposed by Hainmueller, Hopkins, and Yamamoto ([2014](#ref-hainmueller_causal_2014)) that accommodates a vast range of design choices and ensures that they are adequate for drawing accurate inferences. In addition to the derivation of the minimal required sample size given the number of respondents and experimental conditions, our approach adds to the existing literature by providing guidance on the exchangeability between respondents and repeated tasks. In applied research, this can be used to obtain useful insights on whether designs with low number of respondents who fulfil a large number of tasks is as good as having a larger number of respondents who perform just a few tasks. Given a range of pre-specified effect sizes, I also allow researchers to easily adjust the degree of confidence of the estimated parameters and, thus, directly evaluate the trade-off between the false-positive and false-negative. Lastly, the proposed framework accommodates more complex theoretical scenarios or sampling strategies. For instance, effect heterogeneity can be included to perform power calculations when researchers anticipate that two or more groups of respondents have substantially different preferences towards some of the attributes included in the design (e.g., [Kirkland and Coppock 2018](#ref-kirkland_candidate_2018)).

# Validation and Application

The calculation of power analysis for conjoint experiments that use the approach proposed by Hainmueller, Hopkins, and Yamamoto ([2014](#ref-hainmueller_causal_2014)) is done by means of an extensive simulation study. The thesis is based on the working paper of Lukac and Stefanelli ([2020](#ref-lukac_stefanelli_2020)). The aim is to investigate how power changes depending on several features: the sample size, number of tasks performed by each respondent, the number of levels of an attribute, and the size of the measured effect in the population. Furthermore, I will investigate the cases when conjoint experiments are used to discover treatment heterogeneity in two different sub-populations (e.g., female and male). In order to do this, I will simulate data from a conjoint experiment coming from a single, two or more sub-populations with heterogeneous treatment effects and with different prevalence of sub- populations. The simulation will also take into account situations where observations are organized in clusters (e.g., countries) and will investigate the minimum number of respondents needed to benefit from the cluster robust variance matrix estimator’s asymptotic properties ([Arceneaux 2005](#ref-arceneaux_2005_using)).

# References

Arceneaux, Kevin. 2005. “Using Cluster Randomized Field Experiments to Study Voting Behavior.” *The Annals of the American Academy of Political and Social Science* 601 (1): 169–79. <https://doi.org/10.1177/0002716205277804>.

Gelman, Andrew, and John Carlin. 2014. “Beyond Power Calculations: Assessing Type s (Sign) and Type m (Magnitude) Errors.” *Perspectives on Psychological Science* 9 (6): 641–51. <https://doi.org/10.1177/1745691614551642>.

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. “Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments.” *Political Analysis* 22 (1): 1–30. <https://doi.org/10.1093/pan/mpt024>.

Kertzer, Joshua D, Jonathan Renshon, and Keren Yarhi-Milo. 2019. “How Do Observers Assess Resolve?” *British Journal of Political Science*, June, 1–23. <https://doi.org/10.1017/S0007123418000595>.

Kirkland, Patricia A., and Alexander Coppock. 2018. “Candidate Choice Without Party Labels:” *Political Behavior* 40 (3): 571–91. <https://doi.org/10.1007/s11109-017-9414-8>.

Knudsen, Erik, and Mikael Poul Johannesson. 2018. “Beyond the Limits of Survey Experiments: How Conjoint Designs Advance Causal Inference in Political Communication Research.” *Political Communication* 0 (0): 1–13. <https://doi.org/10.1080/10584609.2018.1493009>.

Louviere, Jordan J., David A. Hensher, Joffre D. Swait, and Wiktor Adamowicz. 2000. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511753831>.

Lukac, Martin, and Alberto Stefanelli. 2020. *Conjoint Experiments: Power Analysis Tool*. <https://mblukac.shinyapps.io/conjoints-power-shiny/>.

Maxwell, Scott E. 2004. “The Persistence of Underpowered Studies in Psychological Research: Causes, Consequences, and Remedies.” *Psychological Methods* 9 (2): 147–63. <https://doi.org/10.1037/1082-989x.9.2.147>.

Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716–16. <https://doi.org/10.1126/science.aac4716>.

Orme, Bryan. 1998. “Sample Size Issues for Conjoint Analysis Studies.” *Sawthooth Software Research Paper Series. Squim, WA, USA: Sawthooth Software Inc*. <https://pdfs.semanticscholar.org/a696/a7f303fba91009aaaf1fb2db8c730f39ea05.pdf>.

Rose, John M., and Michiel C. J. Bliemer. 2013. “Sample Size Requirements for Stated Choice Experiments.” *Transportation* 40 (5): 1021–41. <https://doi.org/10.1007/s11116-013-9451-z>.

Schuessler, Julian, and Markus Freitag. 2020. “Power Analysis for Conjoint Experiments.” SocArXiv. <https://doi.org/10.31235/osf.io/9yuhp>.

1. [Alberto Stefanelli](https://albertostefanelli.com/) is an FWO Ph.D Fellow at the Institute for Social and Political Opinion Research and a Fulbright Visiting Fellow at Yale University and New York University. Email: [alberto.stefanelli@kuleuven.be](mailto:alberto.stefanelli@kuleuven.be) [↑](#footnote-ref-20)