Research Statement

Simon Hoellerbauer

My research has two distinct focal points: a methodological one centered on data quality, and a substantive one centered on the political economy of development and civil society. My methodological research focuses on identifying and addressing issues of data quality. There are four main sources for data quality issues – 1) respondents, 2) enumerators/implementers, 3) instruments, and 4) interactions of previous three sources — and my work engages with all four sources. I particularly interested in *assessing* data quality in actionable ways. Our ability to extract causal implications from our analyses relies on the quality of the underlying data, and researchers need to pay more attention to the ways data quality concerns shape our work. Incorporating uncertainty about data quality in our analyses should be a key analytical step. While considerable attention has been paid to the ways instruments affect data quality, less attention has been paid to the other sources of data quality issues. My dissertation project explores two ways in which enumerators and one way in which the interaction between respondents and instruments can impact data quality. Other strands of this track involve the study of measurement - I am particularly interested in developing better measures for civil society and regime types. Finally, another set of projects looks at conjoint survey experiments. In all of these projects, I use a diverse array of methods, including machine learning, Bayesian statistics, and developing novel probabilistic models. My methodological interests directly support my substantive research. Substantively, I study the political economy of development. I am particularly interested in the role that civil society organizations such as NGOs and grassroots organizations play in development and how development support from donors - both private and from other states - affects civil society organizations and civil societies in recipient countries. I use survey experiments, field experiments, and other innovative causal inference methods in these projects.

Dissertation Project

In my dissertation, I tackle three sources of data quality concerns related to surveys and develop methods to incorporate uncertainty about data quality into analysis. In the social sciences, we know that data quality can be an issue, not only with observational data, but also in survey experiments and lab-in-field experiments. Poor quality data can lead to concerns about the generalizability and accuracy of results. We often deal with data quality issues by assuming them away as trivial or by discarding data. My research will be useful to any scientist who collects data – we need to incorporate uncertainty about data quality more directly into our analyses.

In the first paper, "Mixture Models and Survey Quality," I develop a new probabilistic model to diagnose and address data quality issues in cross-sectional and panel data. Researchers can use this model to assess respondent, enumerator, and instrument data quality issues. Using parametric unsupervised machine learning, I estimate uncertainty about survey data in cases where two sets of information exist for respondents, but where there is uncertainty about whether the two sets of information actually correspond to the same individual. In this project, I apply this model to two specific cases: 1) when re-interviews (also called field audits or backchecks) are done (usually deterministically) to assess survey quality, and 2) panel surveys. In the case of re-interviews, the model parameter estimates can be used to assess interviewer performance. The estimation procedure also provides the probability

that an original survey observation represents a real individual or not. In the case of panel surveys, the model allows us to estimate how certain we are that the correct individuals were re-interviewed in a subsequent wave. This is especially critical for in-person surveys in developing countries, where it can be difficult to contact the same individuals. In both cases, we can incorporate uncertainty estimates into our analysis as weights, down-weighting observations of whose accuracy we are not sure. This means we can avoid wasting data. I test the model and its utility using simulation studies built on real-world data from a survey project in Malawi with which I am involved and also via real-world applications.

The second paper, "Implementers as treatment versions: sources and implications of implementer-induced treatment effect heterogeneity," co-authored with Jim Qian (Princeton) and Brandon de la Cuesta (Stanford), presents and tests a new theory of how enumerator-level characteristics affect bias and accuracy in interpreting experimental treatment effects. In this work, we directly engage with the understudied role enumerators can play in data quality. We theorize that different experiment implementers can lead to distinct treatments, beyond those desired by the researcher. We investigate whether certain interviewer characteristics, such as age, experience, and quality, shape enumerator-level treatment effects. Preliminary analysis of several lab in field experiments from Uganda shows that treatment effects are impacted by interviewer characteristics. Given that enumerator characteristics are not randomly assigned, this presents problems for inference. We are developing a more expanded survey of interviewers who worked on a set of surveys for which we have access to the data. In this new survey, we will ask interviewers a series of questions that allow us to assess their quality, experience, and psychological attitudes. We aim to see how systematically interviewer treatment effects vary by these factors and to develop a way of diagnosing this issue.

In the third paper, I examine data quality issues that can arise from a mismatch between respondents and instrument. I test the efficacy and examine the implications for data quality of using images in the presentation of conjoints versus using spoken word vignettes in low-literacy environments. Conjoints are survey experiments in which respondents view randomly created profiles and are asked questions about them. The suggested best practice for doing research with conjoints in low-literacy contexts is to use pictures instead of words when showing profiles to respondents, even when the survey is enumerator-implemented. However, the utility of this strategy has not been empirically tested. There are also additional external validity concerns with an image based conjoint – how does this map onto how respondents would receive similar information in the real world? In this project, I experimentally vary how conjoint profiles are presented to respondents and assess how this affects responses.

Development Research

The core thrust of my substantive research looks at how donor interventions shape civil society by examining when individuals are more likely to engage with organizations. I theorize that individuals are more likely to want to engage with organizations when they are closer to them in a latent values space, which I term congruence, and that individuals take cues from organizational attributes about potential congruence. In the first paper in this research track, "Congruence, Organizational Attributes, and Civic Engagement with Civil Society," which has received a revise and resubmit from the *Journal of Experimental Political Science*, I find support for this theory using a conjoint survey experiment. In the follow-up project, I am testing the theory as directly as possible by developing a novel way

to analyze conjoint survey experiments. I use an IRT model to place organizations and individuals in the same latent space, and then see what effect the distance between them has on engagement using a GLM. I have run an initial pilot study in the United States of America, which finds that students prefer organizations that are closer to them in a latent space defined by organizational attributes.

I am also currently gathering data for a project quantitatively assessing organizational changes caused by donor pressure. The mainly qualitative and case study-oriented development literature has argued that donors have changed how organizations in developing countries operate and present themselves. I seek to test this empirically. Additionally, I am planning a lab-in-field experiment to study how the presentation of information at meetings of a CSO affects subsequent attendance at meetings. This project will test how "NGO-speak" affects citizen engagement with civil society organizations.

Conjoint Survey Experiments

I am pursuing other work on conjoint survey experiments in addition to the paper in my dissertation. Related to my dissertation paper are two projects that expand on the research being done into how the conjoint instrument impacts data quality. In the first, I see if there is a drop-off in data quality the more outcome questions are asked after each profile pair. In the second, I see how expanding the number of profiles seen in a conjoint task impacts data quality. Typical conjoints show respondents two profiles at a time, but in the real world, individuals are not usually presented with only two choices.

Given recent controversy about how to interpret the results of conjoint survey experiments, I am also working on a novel way to analyze conjoint survey experiments that more closely matches an individual's real world decisions making process. In my paper, "Distances in Latent Space: A Novel Approach to Analyzing Conjoints," I propose a two-part model that combines an IRT model with a GLM. I use two outcomes, one to put individuals and profiles into the same latent space, and the second to assess how the distance between individuals and profiles impacts behaviors. I have run a pilot study on a student sample and will run a second study during the fall 2021 semester.

Similarly motivated by the need to re-examine how we analyze conjoints, I plan to apply the logic behind ANOVA to clarify inference around conjoint attributes in conjoint survey experiments. Researchers mostly look at attribute-levels when employing conjoints, but sometimes we are agnostic about the impact of attribute-levels and are actually hypothesizing about the importance of whole attributes.

Measurement

The ways in which we measure concepts can impact data quality as well - if we are capturing the concept we think we are, it has serious implications for our analysis. I am working on a project, "Reconceptualizing Civil Society and its Strength," in which I argue that many of the ways in which researchers have measured civil society strength, both qualitatively and quantitatively, suffer from endogeneity issues and misunderstand how civil society works. I develop a theoretical model for measuring civil society strength. The approach I lay out can be used by quantitative and qualitative researches when assessing civil society strength.

In another project, "Democracy vs Dictatorship or Something More?: Using Unsupervised Learning to Cluster Regimes," I propose using machine learning – in particular unsupervised or semi-supervised clustering – to create regime typologies. In political science, we have many different categorical typologies of regime type. While such typologies are very

useful for understanding and studying regimes, deciding where states fall on these typologies can involve many arbitrary decisions. We can see what regime types naturally occur in the data using machine learning.

Other Current and Future Methods Research

I also am working on a project using double/debiased machine learning to make marginal structural models more flexible with Santiago Olivella (UNC-Chapel Hill) and Matt Blackwell (Harvard). Marginal structural models allow researchers to limit post-treatment bias when there are time-varying confounders. We combine marginal structural models with machine learning. Because machine learning generally sacrifices bias for the sake of variance reduction, we use debiased or double machine learning to estimate causal quantities of interest.

I am also in the early stages of projects that seek to expand existing methodological approaches and bring cutting-edge statistical methods to political science. In one, I re-examine the mainly visual use of predicted quantities in political science and propose clearer ways to do inference using predicted quantity plots. In a third, I aim to explain double machine learning and its utility for questions of interest in political science.

Other Current Development Research

I am also collaborating with Lucy Martin (UNC-Chapel Hill), Brigitte Seim (UNC-Chapel Hill), and Luis Camacho (NORC) on a project, "Marketing Taxation? Experimental Evidence on Enforcement and Bargaining in Malawian Markets," that analyzes the impact of a USAID-funded intervention in Malawi. The intervention tested whether citizen or state empowering strategies were more successful at increasing tax compliance in Malawian markets. We have workshopped the paper extensively including via EGAP and at APSA 2020 and are in the process of finalizing the paper for journal submission. With Lucy Martin and Brigitte Seim, I am also working on a paper titled "Investigating Tax Policy Preferences Among Citizens in Weak-State Contexts," that analyzes Malawian citizen attitudes toward tax plans using a conjoint survey experiment. We test whether the kinds of taxes citizens support seem to be different from the kind of taxes citizens report being willing to pay.