Observational Studies 2 (2016) 86-89 Submitted 10/16; Published 10/16

Book review of "Causality in a Social World" by Guanglei Hong

Kenneth A. Frank **kenfrank@msu.edu**
Department of Counseling, Educational Psychology and Special Education, College of Education
Department of Fisheries and Wildlife, College of Agriculture & Natural Resources
Michigan State University
Lansing, MI, U.S.A.

Guan Kung Saw **guan.saw@utsa.edu**
Department of Educational Psychology
College of Education and Human Development
The University of Texas at San Antonio
San Antonio, TX, U.S.A.

Ran Xu **ranxu@msu.edu**
Department of Counseling, Educational Psychology and Special Education
College of Education
Michigan State University
Lansing, MI, U.S.A.

As the introduction of Guanglei Hong's *Causality in a Social World* makes clear, this book would not be necessary if all treatments we wished to study had constant effects through simple mechanisms on independent individuals who were randomly assigned to treatments. While, such conditions may hold in some idealized agricultural settings, this is not the phenomenon we encounter in a social policy oriented world with human agency. In response, Hong presents a coherent theoretical and empirical framework for estimating causality when people choose their own treatments, when they encounter mediating and moderating effects of treatments and when they influence others' choices and outcomes. The book is presented in four large sections: overview, moderation, mediation and spill-over, with a chapter introducing the core ideas in each section (chapters 4, 7, 11 and 14 respectively). Beyond merely consolidating her own foundational work, the book is steeped in deep and historical statistical principles of sampling, propensity score analysis, mediation and moderation, and spill-over mechanisms. Ultimately, the book will mark a passageway from underlying statistical principles to a framework that may endure and expand beyond even what Hong anticipates.

The core of the framework is to conceptualize causal inference as a sampling issue, building upon a long tradition of statistical adjustment through weighting in survey sampling. In particular, Chapter 4 lays a foundation in conventional statistical principles, in this case of sampling, stratified sampling, and weights to correct for purposeful sampling that create a sample disproportionate to population treatment assignments. The turn then is to apply the sampling conceptualization to strata defined by the propensity to have been assigned to the treatment.

When subjects are placed in strata based on their propensity to receive the treatment the proportion of treated and control subjects within each stratum will typically not represent their respective proportions in the full sample. This creates challenges in synthesizing results across strata. The insight is that the data within each stratum can be weighted to reflect the proportions of treated and control subjects in the overall sample (or in the population if people are randomly sampled). This is known as marginal mean weighting through stratification (MMWS). Drawing on her own seminal work in this area, Hong shows that for binary treatment estimates based on MMWS are equivalent to estimation through stratification without weights. Furthermore, both outperform (in terms of root mean square error – page 94) conventional inverse proportional treatment weighting (IPTW). This is in part because assignment to discrete strata is robust to misspecification of the propensity model relative to continuous weighting approaches (see page 98), while the weights insure that the treatment and control cases in each stratum contribute proportionally to the overall estimate.

In addition to its estimation advantages, MMWS is more flexible for studying binary and multivalued treatments than conventional propensity score matching and stratification that are mostly limited to examinations of dichotomous treatments. Extensions to mediation, moderation and spill-over are based on a conceptualization of multiple treatments, either constructing a sequence as in mediation, separable regimes as in moderation, or a two-stage clustered randomized design as in spill-over.

Just as chapter 4 presents MMWS within traditional sampling ideas, so do each of the parts show how new ideas emerge out of existing techniques. For example, chapter 6 presents moderation in the context of randomized experiments and factorial designs, chapter 9 builds mediation out of Baron and Kennys (1986) path diagrams, and chapter 14 builds on some of Hong's (and Raudenbush's) original work to integrate standard models network effects (equation 14.4) into Rubin's potential outcomes framework. In this sense the book is an excellent introduction to several key statistical concepts in the social sciences, integrated through the enterprise of inferring cause in social context through a sampling lens.

The book's solid scholarship allows one to understand how newer techniques can generate conventional results under restricted assumptions. For example, chapter 4 describes how marginal mean weighting reproduces IPTW when covariates are discrete and chapter 11 shows how the Ratio of Mediator Probability Weighting (RMPW) reproduces path analysis and instrumental variables approaches when there are no interactions between mediator and treatment or when the exclusion restriction holds (there is zero direct effect from treatment to the outcome). This helps researchers understand in what ways the new techniques are improvements rather than mere fads, thus contributing to scientific accumulation (although at times conventional techniques are not always presented as they are typically applied weights are typically trimmed in IPTW to make estimates more efficient).

While this book offers thoughtful synthesis and advancements it is important to understand the limitations. In particular, the MMWS method is grounded in propensity score approaches. Such propensity scores are only as good as the covariates used to define the propensity model propensity scores do not inherently account for bias due to variables omitted from the propensity models (Rosenbaum, 2002, page 297; Shadish et al., 2002, page 164; Heckman, 2005; Morgan and Harding, 2006, page 40). Hong knows this well, as she deliberately and consistently presents and invokes the strong ignorability assumption that treatment assignment is independent of potential outcomes once conditioning on covariates in the model (see chapters 2-7). That is, absent a treatment effect there is no expected difference between treatment and control groups on the outcome that is a function of omitted variables. Nonetheless it is critical that readers be aware of this limitation as the methodological advancements in the book depend heavily on propensity scores but the emphasis is on what is done with the propensity scores more than on the specification of the propensity model itself. Moreover, Hong helps the reader engage concerns about misspecification by presenting sensitivity analyses that quantify the robustness of inferences to misspecification and generally to bias (see chapters 3 and 10).

Beyond the assumption of strong ignorability, there are other assumptions of propensity models that could be more explicit. First, propensity scores themselves are only estimates. As such, variability in their estimation should be accounted for in terms of the degrees of freedom or by bootstrapping the process of estimation of the propensity model and the final model (see review in Caliendo and Kopeinig, 2008). Although use of bootstrapping when matching cases is not uniformly accepted (Abadie and Imbens, 2008), it has more potential with stratification techniques because they operate on large groups of discretely defined samples. Second, relatively large samples are required to populate both the treatment and control cells within each stratum in an MMWS or an RMPW approach. The National Evaluation of Welfare to Work Strategies (NEWWS) example in chapter 11 has roughly 700 cases, which when broken down by treatment within strata could lead to small cell sizes.

Pedagogically, the strength of this book is in taking the reader from what is already known to the new in a very deliberate and clear way, making Hong's and others' insights accessible and applicable. There are small pedagogical limitations. At times Hong emphasizes the conceptual and general formulas over an immediate and fully integrated empirical example (chapter 4). In this sense the book is intended for the reader well versed in basic statistics (e.g., the general linear model, p-values) who is ready to engage a new, well-grounded, flexible conceptualization to estimate cause in the complex social world attended to by policy research. Such a reader would find the entire book, from core chapters to extensions of techniques to technical appendices accessible and valuable. And there is a large pay-off for the reader who invests in this book both in terms of improved estimation of treatment effects as well as a new, integrated perspective on some traditional techniques.

References

Abadie, A. and Imbens, G.W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, 76, 15371557.

Baron, R.M. and Kenny, D.A. (1986). The moderatormediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22, 3172.

Heckman, J. (2005). The scientific model of causality. *Sociological Methodology*, 35, 199.

Morgan, S.L. and Harding, D.J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods and Research*, 35, 360.

Rosenbaum, P. (2002). *Observational Studies*, 2nd edition. Springer: New York.
Shadish, W.R., Cook, T.D., and Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin: Boston.