Causal inference in travel demand modeling (and the lack thereof)

Timothy Brathwaite [a],[*], Joan L. Walker [b]

[a] Department of Civil and Environmental Engineering, University of California, 116 McLaughlin Hall, Berkeley, CA, 94720-1720, USA
[b] Department of Civil and Environmental Engineering, University of California, 111 McLaughlin Hall, Berkeley, CA, 94720-1720, USA

A R T I C L E   I N F O

A B S T R A C T

This paper is about the general disconnect that we see, both in practice and in literature, between the disciplines of travel demand modeling and causal inference. In this paper, we assert that travel demand modeling should be one of the many fields that focuses on the production of valid causal inferences, and we hypothesize about reasons for the current disconnect between the two bodies of research. Furthermore, we explore the potential benefits of uniting these two disciplines. We consider what travel demand modeling can gain from greater incorporation of techniques and perspectives from the causal inference literatures, and we briefly discuss what the causal inference literature might gain from the work of travel demand modelers. In this paper, we do not attempt to "solve" issues related to the drawing of causal inferences from travel demand models. Instead, we hope to spark a larger discussion both within and between the travel demand modeling and causal inference literatures. In particular, we hope to incite discussion about the necessity of drawing causal inferences in travel demand applications and the methods by which one might credibly do so.

1. What demand modelers have always wanted to do

Consider the following three quotations.

"Travel demand models are used to aid in the evaluation of alternative policies. The purpose of the models is to predict the consequences of alternative policies or plans. […] Predictions made by the model are conditional on the correctness of the behavioral assumptions and, therefore, are no more valid than the behavioral assumptions on which the model is based. A model can duplicate the data perfectly, but may serve no useful purpose for prediction if it represents erroneous behavioral assumptions. For example, consider a policy that will drastically change present conditions. In this case the future may not resemble the present, and simple extrapolation from present data can result in significant errors. However, if the behavioral assumptions of the model are well captured, the model is then valid under radically different conditions." —(Ben-Akiva, 1973)

"Indeed, causal models (assuming they are valid) are much more informative than probability models. A joint distribution tells us how probable events are and how probabilities would change with subsequent observations, but a causal model also tells us how these probabilities would change as a result of external interventions—such as those encountered in policy analysis, treatment management, or planning everyday activity. Such changes cannot be deduced from a joint distribution, even if fully specified." —(Pearl, 2009b)

"The goal of many sciences is to understand the mechanisms by which variables came to take on the values they have (that is, to find a generative model), and to predict what the values of those variables would be if the naturally occurring mechanisms were subject to outside manipulations. [...] Finding answers to questions about the mechanisms by which variables come to take on values, or predicting the value of a variable after some other variable has been manipulated, is characteristic of causal inference."—(Spirtes, 2010)

Based on personal communication with many travel demand modelers, i.e. based on anecdote, we believe that the first quotation, by Moshe Ben-Akiva, accurately represents the opinions of most researchers and practitioners within the field of transportation. Moreover, we think it is safe to say that a "policy that will drastically change present conditions" can be categorized as an "external intervention" or "outside manipulation." If one accepts these two premises, then based on the two quotations by Pearl and Spirtes, it is clear that the implicit goal of travel demand modeling is to make causal inferences (i.e. "to predict the consequences of alternative policies or plans")[1]. Moreover, in order to produce such causal inferences, it is clear that travel demand models should be "causal models."

In the rest of this paper, we further investigate the relationship between travel demand models and "causal models" as seen in other disciplines. Section 2 provides a brief overview of what causal inference is and why it should be seen as a distinct field from travel demand modeling. In Section 3, we describe the current state of relations between the fields of causal inference and travel demand modeling. There, we pay special attention to the differences between practices in the causal inference literature and practices in travel demand modeling. In Section 4, we continue this focus by hypothesizing about why the travel demand modeling literature seems so far removed from the causal inference literature. Finally, although we do not try to "solve" the issues of drawing valid causal inferences from travel demand models, we try to bridge the gap between the two literatures in Section 5. In this section, we emphasize what travel demand modelers can learn from causal inference researchers, we provide an extended example that illustrates the use of the techniques described in this paper, and we conclude with a statement about how travel demand modelers can contribute to the causal inference literature.

2. A brief primer

Despite sharing the same goals (as highlighted in Section 1), we believe travel demand modelers are generally uninformed or misinformed[2] about key concepts from the field of causal inference.

Here are some recent examples of this point. On April 20th and 21st, 2017, the "Advancing the Science of Travel Demand Modeling" National Science Foundation Workshop was held at the University of California, Berkeley. This workshop convened many travel demand modeling scholars and practitioners, young and old, from both within and outside of the United States. As such, the comments made during the workshop represent a wide cross-section of voices within the field. Of special interest was panel discussion #2: "How critical is causality? And how can we make clear statements about causality in travel demand models?" In particular, some direct quotes[3] from the discussion after Panel #2 were:

- "What is causality? What is the clear definition of causality?"
- "What is causality? What about the context? It's not just Y and X."
- "How do we define causality? How much causality is needed in the models to give robust predictions?"
- "A model that predicts successfully implies that we are accounting for causality."
- "If we take a certain intervention, will it have the outcome desired by the policy makers? It's not about getting causality right. It's more about what confidence do we have in our projected outcome."

As illustrated by these comments from attendees and the overall tenor of the conversations throughout the workshop, the topic of causality in travel demand modeling is beginning to be widely discussed, but it is still far from being widely and correctly understood[4]. Specifically, travel demand modelers seemed most uninformed or misinformed about what causal inference is and how it differs from prevailing practices in travel demand modeling. Below, we briefly address these two questions.

First, for the purposes of this paper, causal inference is defined as the use of data and assumptions to make inferences about outcomes

---

[1] As noted by an anonymous referee, "some might argue that the purpose of demand modeling is to make predictions, as opposed to discover the causal mechanism." We believe such distinctions are red herrings. The predominant role of travel demand modelers, especially practitioners, is to predict the effects of particular policies on a future population's travel behavior. As stated by Spirtes (2010), "predicting the value of a variable [i.e. travel behavior] after some other variable [i.e. a policy] has been manipulated is characteristic of causal inference." Put succinctly, counterfactual prediction is a causal inference task. Identifying causal mechanisms is also a causal inference task, but identifying causal mechanisms is not always necessary for making counterfactual predictions.

[2] Note, we do not use the adjectives "uninformed" and "misinformed" to be disparaging. We mean very literally that travel demand modelers do not seem to widely read the causal inference literature, and because the concepts and findings of that literature are non-trivial and sometimes un-intuitive, travel demand modelers often express sentiments that (1) show a lack of awareness of the technical details and definitions from the causal inference literature or (2) show beliefs that directly contradict findings from the causal inference literature. This second point is supported in the next paragraph.

[3] Note that the names of individuals who made each quote have been redacted to respect participant privacy because individuals did not make these statements "on the record."

[4] To be completely explicit, we note that on the topic of making inferences about outcomes under external manipulation or intervention, we generally assume that if the statements of travel demand modelers and causal inference researchers disagree, then the travel demand modeler is incorrect. Of course, we examine the statements and supporting arguments made by both parties, but we have found our assumption to typically hold true. Again, this is not a pejorative remark against travel demand modelers. It is an expected outcome based on the fact that causal inference researchers are trained to focus on this topic, whereas travel demand modelers are typically not.