The Promises and Pitfalls of Using Panel Data to Understand Individual Belief Change

Revision Memo

December 2024

We would like to thank the reviewers and the editors for their excellent and thorough comments. We especially appreciated the reviewers' clear suggestions that allowed us to revise major portions of the first manuscript. In the following pages, we outline the changes we made in the paper.

1 Substantive Issues

1.1 The Take-Home Message of the Article

Both the reviewers and the editors urged us to clarify the take-home message of our article.

Particularly, R1 noted that:

> Readers may be uncertain of how to use the authors' findings.

Similarly, R3 noted that:

> I remained puzzled if this [measurement error] is an issue the proposed model solves or the authors are just conscious of its existence.

We appreciated this feedback and made substantial rewrites in the body of our manuscript to clarify what the goal of the analyses is and why researchers should implement our method.

In a nutshell, we would like to note that we do not claim to have solved the measurement error problem. Rather, our aim is much more modest: we want to provide applied researchers a tool that will allow them to see what plausible DGPs may have generated their data. This might be thought of as something akin to a power analysis for theoretical claims. In the context of this special issue, this means that we urge researchers to think more carefully about how different compositional claims—a combination of rate of change, strength of change, and the mix of directional change—may be responsible for the same data, such that our claims about individual preferences might have different implications depending on different DGPs.

As we detail below, we now provide an R package that researchers can seamlessly implement in their own work to perform this analysis, which we hope will clarify our message.

1.2 The Discussion of the Group-to-Person Generalizability Problem

Reviewers suggested that an extended evaluation of the group-to-person generalizability problem is warranted in the manuscript to highlight the importance of the measurement problem.

Particularly, R1 noted that:

> Examples of recent or important theories of individual-level change might highlight the tension between the pervasiveness of this kind of theorizing with the measurement challenge outlined in the paper.

Similarly, R3 noted that:

> I would encourage the authors to formalize the issue already in the introduction and highlight it with examples.

We extended our discussions with these suggestions. Particularly, we now walk through an example of how average coefficients represent a combination of our three quantities of interest.

1.3 The Item Composition

R1 notes that

> Do many—or many politically important—items look like the marijuana legalization item or the abortion item?

Similarly, the editor notes that

> The panel data were collected over a relatively short time span. Does this matter? It may possibly make it harder to detect belief change.

We agree with these concerns. We did two changes to improve upon these points:

- We added a new item using the *British Elections Study*'s panel study, generalized trust, an item collected between 2019 and 2024 across 6 years and 5 waves, in a potentially disruptive time period due to COVID-19. As a more general note, we believe that the tool we propose should apply in all cases, but we agree that a more thorough set of examples is helpful.
- Additionally, we now present a simulation study that applies our protocol in different DGPs and assesses the conditions under which we can make better guesses about the true DGPs (see below for more details). We think this new section provides important answers about the questions of generalizability, panel span, varying DGP compositions, and related issues.

2 Issues on Analytic Strategy and Implementation

2.1 The Grid-Search Procedure

R3 notes that

> "the five bullet point explanation of the Grid-Search Procedure is inadequate."

We implemented three solutions and a larger revision for this problem:

1. We updated the mechanics of our approach by using Approximate Bayesian Computation (Beaumont 2010), which increased the efficiency and generalizability of our procedure.
2. We now introduce our grid-search algorithm as an accompanying R package, `gridsearch`. As we detailed in the paper, we revised our procedure and substantially increased its efficiency, which means that it will be easy for people to implement our method in their own projects.
3. We included a full (anonymized) replication package for the analyses presented in this project.
4. We made substantial rewrites to the manuscript to increase the clarity of our protocol.

We hope that the R package, the replication repo, and these rewrites make our strategy more clear.

2.2 Simulations and Validations

As R2 noted, our grid search protocol will perform better in some cases than others—e.g., having few people changing their beliefs in the panel context might make it hard to find a plausible DGP under low reliability, or having a mixture of positive and negative changes might complicate our search approach. To address these issues, we conducted two simulation studies, as we described in a new section in the manuscript. These simulation studies allowed us to validate our procedure, while at the same time showing the conditions under which we may expect high uncertainty from the data. These studies strengthened our paper greatly and we thank R2 for their great suggestion.

2.3 The Detection of True Changers in the Sample

R2 notes that

> "One thing I was concerned about in the end is that the third question—who changed—seems not have been responded to adequately."

We agree. That said, we believe that the reason for this is *substantive* rather than *methodological*. As stated above, our goal is to provide researchers a general tool to assess plausible DGPs and the conditions under which they can make claims about (1) compositional distributions such as rate of change and strength of change and (2) their own ability to detect "true changers." We are less interested in providing particular solutions to these problems, as they are context-specific.

This is particularly important when it comes to R3's comments on the estimator choice:

> "The solutions I am familiar with to the heterogeneity of effects problem involve mixture models that allow the different mixtures to highlight different effects in population sub-groups [...] if the proposed model works better than these, I would have expected to see explicit comparisons or at least direct contrasts. Such models are only passingly mentioned in the conclusion. A fixed effects model explain these away entirely while the two modeling approaches discarded allow for their modeling."

We agree. This issue is not a concern for the grid search algorithm, where our central objective is to generate a distance score between real data and simulated data, but it is indeed a concern when it comes to providing classification scores for whether we can recover *true changers*. As R3 stated, the use of varying individual fixed effects is one choice among many, and there are many alternatives.

While we agree with this point, we have two reasons for refraining from these extensions. First, these models proved problematic when it comes to scaling. Remember that we perform N simulations for detecting true changers, and map the true changers we know from the DGP to the observed ones we recovered from our estimation. In the context of population-wide growth curve models, partial pooling causes a strong shrinkage in the estimates that, in turn, increases the researcher degrees of freedom as to how we can classify individuals—this is not a concern in the fixed-effects case, where the estimates have a basic decision rule with a benchmark using the strength parameter in the DGP. This is easier to circumvent in the context of mixture models, where the aim is to classify in the first place. That said, we found that the choice of model parameters affects these classifications in panels with low number of Ts, which is why we decided to stick to the fixed-effects method.

We now have an extended discussion about these issues in the `Discussion` section and we provide specific suggestions for further research in this area. We thank R3 for their keen observation.

3 Issues on Presentation and Results

3.1 The Size of Changes

R1 notes:

> I believe the authors are referring to change in latent positions, but this should be made explicit.

That is true. We always separate the latent positions and the observed survey positions. However, we agree that the discussion of the change scores was not sufficiently clear in the manuscript. We thus changed the relevant sections to emphasize this difference. We thank the R1 for this warning.

3.2 Reliability and Resolution

R1 notes:

> Varying parameters relating to reliability and resolution in simulations might be useful, given the prominence of these things in the conclusion section.

We have two comments regarding this suggestion:

(1) *Reliability*: We agree with R1 that an exploration of reliability is warranted. To provide a more complete picture of how "successful" we can be as a function of different reliability scores, we conducted simulation analyses. The manuscript now includes a section on how real reliability and assumed reliability in the grid search interact to produce different error values.

(2) *Resolution*: One important weakness of our current manuscript is that we restrict our attention to the binary cases. The generalization of our procedure to cases with more than two response categories, however, requires a substantial addition to our protocol—an extension that we are currently working. We included an extensive discussion about this issue in the `Discussion` section, urging future research about this issue. We thank R1 for their suggestion.

3.3 The Grid Search Plots

The editor notes:

> I would have also liked a bit more explanation on how to read the fake data simulation graphs.

In response to this, we revised our plots and extended our discussions about how to interpret them.

Thank you all for this intellectual exchange, and we greatly appreciate your deep engagement with our article. We hope that the revisions we made adequately responded to your concerns.

Regards,

The Authors

References

Beaumont, Mark A. 2010. "Approximate Bayesian Computation in Evolution and Ecology." *Annual Review of Ecology, Evolution, and Systematics* 41(1):379–406.