

Jeff Gill

Editor-in-Chief. Email: jgill@american.edu

Welcome to the first *Political Analysis* replication forum. This was not a planned event, but something that grew organically, and by necessity, over the last 3 years. In 2016 *Political Analysis* published an important and influential article by Muchlinski, Siroky, He, and Kocher advocating random forest models over conventional logit regression to predict the onset of civil war. Subsequently in 2017 two separate and completely independent Letters (Neunhoeffer and Sternberg 2018; Wang 2018) were submitted after finding the technical issues of estimation and prediction discussed in this forum. These manuscripts were reviewed and are contained herein. This also commenced a detailed replication and review process within the editorial team at *Political Analysis* that produced an additional Letter (Heuberger 2018) focusing on technical details of the code and the process. Finally, after much discussion internally and externally we decided that the best outcome would be a positive jointly produced set of Letters in a forum that not only discusses the original analysis of Muchlinski *et al.* (2016), but also the difficulties in replicating complex, sophisticated work in the current era. We add to this chronology another letter received later in the process that independently discusses issues, challenges, and prescriptive advice in replication. Jeffrey Hardin, Anand Sokhey, and Hannah Wilson develop a framework for improving replication and apply it to a preregistered replication study. Our hope is that readers will not only enjoy reading about the issues, and chronology in these works, but will also fully appreciate some challenges we face in evaluating the work of our peers in political methodology.

In my view the most important takeaway from the process that created this forum is the value of the *Political Analysis* replication process put in place by my predecessors R. Michael Alvarez and Jonathan Katz. What it means is that we as a subfield evaluate and quality-control ourselves, catching issues that need to be addressed so that scholars can use the knowledge, methods, and approaches published in *Political Analysis* with confidence that they are not just vetted at review time but also throughout the life of the article. I know of no other subfield of political science that is as intensively self-critical and self-reflective in this way. I hope this forum provides readers of *Political Analysis* with an assurance that we value the reliability of published findings at the highest possible levels.

In 1995 Gary King wrote: “As virtually every good methodology text explains, *the only way to understand and evaluate an empirical analysis fully is to know the exact process by which the data were generated and the analysis produced*” (italics in the original). This remains true today, but not uniformly appreciated *23 years later* (italics are mine here). The community of empirical political scientists has benefited immeasurably by the recent trend in required replication that is now standard at leading journals of the field, and started here at *Political Analysis*. I am the last person to have “natural science envy” as I also pay attention to the big challenges in data analysis in other fields, but *the* standard for acceptance of results in many of these fields is replication where failed replication (or unavailable replication) constitutes nonbelievability. Consider the episode in 2011 where researchers at CERN in Switzerland appeared to have measured a neutrino traveling at faster than the speed of light. They, not having full belief in their own findings even after inspecting them closely, immediately released the data to all of the physics world for confirmation or refutation. Predictably the results failed to be supported, and this was due to a reason that

***Political Analysis* (2019)**
vol. 27:98–100
DOI: 10.1017/pan.2018.44

Corresponding author
Jeff Gill

© The Author(s) 2018. Published by Cambridge University Press on behalf of the Society for Political Methodology.

all readers of this journal understand: statistical uncertainty. See Fargion and D'Armiento (2012). The important part of this story is that data-analytical results in physics are always subject to replication with unambiguous confirmation or refutation. Other fields, like empirical political science are clearly moving in that direction.

At the core of replication is the ability for scholars who are remote from the intricacies of the individual study to gain confidence in the procedure independent of the review process, which typically does not include replication at that stage. There is a substantial set of research decisions that may not make it into the text of the published article: which version of the data were used, recodings of variables, subsetting decisions, handling of missing data, software intricacies, different versions of particular statistics or procedures, validation methods, rounding(!), and more. An interesting experiment would be to give two political scientists the exact same dataset and a single research question and see how different the empirical results are. Here is another interesting experiment you can do at home (moderately senior scholars only): try and replicating your own earliest empirical publication. It is harder than one might expect. So the more we understand about the mechanics of the experiment, the observational data analysis, or even the derivations, the better we progress as a subfield.

What did we learn from this particular episode with the replication of Muchlinski *et al.* (2016). First that journal editorial staff replication alone can be insufficient. The analytical work in R was replicated in-house and the errors were not discovered. This is another reason why it is important to archive the data and code somewhere in a permanent archive such as `dataverse` so that others have access to it. Second, there is an engaged community out there that vitally cares about what appears in the journal and the process by which these articles are created. This is a blessing and moving to an institutionalized replication process has helped cement the relationship. Evidence of this was two independent replications submitted as Letters to the journal for review reanalyzing Muchlinski *et al.* (2016). Third, that the process can be immensely time-consuming. Due to the analytical complexity and the number of authors of different pieces involved, this episode used a lot of editorial time. We first needed to figure out in-house if there was a problem, then get two Letters through the peer review process, then prepare a replication of our own, and of all this including communicating with the original authors. This leads to a final point: it is vital to keep open, positive, and professional communication with all parties rather than make it an adversarial process, which is a human temptation.

The editors at *Political Analysis* thank Marcel Neunhoeffler, Sebastian Sternberg, and Yu Wang, for discovering issues in the original Muchlinski *et al.* (2016) paper, producing insightful Letters, and working to survive a full peer review process. We also thank our Editorial Assistant Simon Heuberger, who as the official replicator at *Political Analysis* was automatically drawn into this process with the editorial staff, and produced the reviewed Letter contained in this forum with the most detailed technical replication issues.

Finally, I would like to thank the cooperation, professionalism, and forthrightness of David Muchlinski, David Siroky, Jingrui He, and Matthew Kocher. Their assistance in getting this forum to publication and their willingness to share their process was invaluable. They should be commended.

References

- Fargion, Daniele, and Daniele D'Armiento. 2012. Inconsistency in super-luminal CERN–OPERA neutrino speed with the observed SN1987A burst and neutrino mixing for any imaginary neutrino mass. *Journal of Physics G: Nuclear and Particle Physics* 39(8): 85002.
- Heuberger, Simon. 2018. Replication analysis of Muchlinski *et al.* (2016). *Political Analysis*, Forthcoming.
- King, Gary. 1995. Replication, replication. *PS: Political Science and Politics* 28(3):541–559.
- Muchlinski, D., D. Siroky, J. He, and M. Kocher. 2016. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Political Analysis* 24(1):87–103.

- Neunhoeffer, Marcel, and Sebastian Sternberg. 2018. How cross-validation can go wrong and what to do about it. *Political Analysis*, Forthcoming.
- Wang, Yu. 2018. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data: A comment. *Political Analysis*, Forthcoming.