Critical Finance Review, 2019, 8: 301–304

Reproducing, Extending, Updating, Replicating, Reexamining, and Reconciling

Ivo Welch

Anderson Graduate School of Management at UCLA, USA; ivo.welch@ucla.edu

There are various forms of reanalysis. I am adopting ideas from Hamermesh (2007), Clements (2017), and Christensen and Miguel (2018) to suggest a basic taxonomy:

- Reanalyses may or may not use the same computer logic (code) as the original paper.
- Reanalyses may or may not use the same sample, and may or may not draw from the exact same population.

These two dimensions suggest at least four categories: reproductions, extensions (including updates), replications, and reexaminations.

1 Reproduction

> **Reproduction: An attempt to replicate the same result in the same sample with the same code.**

Reproductions should be perfectly objective and leave no room for disagreement. Non-identical and merely similar results count as failures in a reproduction. A reproduction can yield different results if the original data sample and computer programs are no longer available and have to be reconstructed. They can fail due to code indeterminacies (e.g., multiple local optima), code changes and corrections, code bugs, and data updates or changes.

Ideally, all published papers should be perfectly and easily reproducible by any skilled scientist. They should be the starting point for related future work. Of course, from a collective perspective, it is not necessary that every reader conducts her own reproduction. It is sufficient that a reasonable number of *skeptical* authors have done so. In real life, this may not always be possible, especially when data and code were not archived or were not archivable, perhaps because they are proprietary.

Reproductions are the most essential type of reanalysis. They are the foundation of science. They are the evidence instead of someone's word. I am going

ISSN 2164-5744; DOI 10.1561/104.00000082

even further. What matters in science is not replicability. What matters is actual replication.

2 Extensions and Updates

> **Extension: An attempt to run the same code on the same population, but with additional sample draws.**
>
> **Update**: A special case of an extension in which the sample is continued in time from where the original study had left off.

An extension can yield different results due to sampling variation, structural changes, or inadvertent population differences. This is especially common in financial markets, where financial market participants likely change their behaviors in response to academic research (McLean and Pontiff, 2016).

The authors of original papers should never have cause to object to the results of pure extensions and updates. After all, extended results are only the content that they would have published themselves if they had written their articles years later. The original authors should also never be blamed if updates change the inference. After all, causality is always local. Except for tautologies or near-tautologies—i.e., the most trivially obvious economic hypotheses which are bound to hold universally (Welch, 2016)—a hypothesis may well hold in one subset of data but not in another.

In extensions and especially in updates, the prescription is simple:

- If past results no longer hold after the data has been updated, the original results have become obsolete.
- If past results no longer hold in the new data alone, it could be either that original results have become obsolete or that the new data alone does not yet have enough power. This can be assessed with common statistical tools.

3 Replication

Disagreements (and sometimes hurt feelings) between original and reanalyzing authors are more likely to arise in the next two types of reanalyses:

> **Replication: An attempt to run the same code on a similar but different population.**

Replications can yield different results due to sampling variation, differences in populations (and thus samples), or structural changes. In empirical economics, it

is rarely the case that we know whether failure to obtain similar results is because of sampling variation (a different sample from the same population), or because of a different population. Thus, distinguishing between extensions (even including updates) and replications is often more aspirational than real.

Replications are one reason why it is important to preserve code and data. The best starting point for a replication is always a reproduction.

However, it is also important that the results can be replicated from an inexact description of the supposedly essential ingredients *without* code and data. Ideally, this description should be separate from the paper's description of its inference in order not to pre-bias the re-examiner.

If a study is replicable (reproducible) only if the methods are followed precisely to the letter using the original code and data, then the original results are not robust. A "near reproduction" replication should not yield the exact same but quantitatively similar results. Put differently, a functional plot of results with respect to unimportant and reasonable variations should not only not have the reported results sit at the maximum, but the function itself should also not be too steep. If this is not the case, how economically meaningful is the broader economic inference?

4 Reexamination

> **Reexamination: An attempt to run extended code on the same sample.**

The "same sample" here means the same sample observations, though not necessarily the same variables. The reexamination can yield different results if it includes a previously omitted important variable or if it uses different econometric analysis tools.

5 Syntheses

The boundaries between the four categories are often gray. Most reanalyses include combinations of all of them. If the original study and a reanalysis come to different conclusions, ideally the reanalysis should offer a **reconciliation**. Again, this can only be done if the reanalysis can start with a reproduction.

Reconciliations help us understand why results are different. Syntheses are also often subject to disagreement. This is not necessarily bad, because these disagreements can shed even more light about how the world works. The CFR has published many papers with both excellent critiques and excellent responses.

6 Conclusion

I strongly endorse the call to arms in Harvey (2017): If financial economics wants to remain credible, our profession must embrace more reanalyses of all types. We must collectively demand that our journals publish many more skeptical reanalyses and we must give more credit for reanalyses in our tenure and promotion processes. More credit is not a cure-all, but it would be a good start.

References

Christensen, G. S. and E. Miguel. 2018. "Transparency, Reproducibility, and the Credibility of Economics Research." *Journal of Economic Literature*. 56: 920–980.

Clements, M. A. 2017. "The Meaning of Failed Replications: A Review and Proposal." *Journal of Economic Surveys*. 31: 326–342.

Hamermesh, D. S. 2007. "Viewpoint: Replication in Economics." *Canadian Journal of Economics*. 40: 715–733.

Harvey, C. R. 2017. "Presidential Address: The Scientific Outlook in Financial Economics." *The Journal of Finance*. 72: 1399–1440.

McLean, R. D. and J. Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance*. 71: 5–32.

Welch, I. 2016. "Plausibility: A Fair and Balanced View of 30 Years of Progress in Ecologics." *Foundations and Trends in Accounting*. 10: 376–412.