



**EJW**

ECON JOURNAL WATCH  
Scholarly Comments on  
Academic Economics

ECON JOURNAL WATCH 11(1)  
January 2014: 11-16

# One Swallow Doesn't Make a Summer: Reply to Kataria

Zacharias Maniadis<sup>1</sup>, Fabio Tufano<sup>2</sup>, and John A. List<sup>3</sup>

[LINK TO ABSTRACT](#)

In his comment, Mitesh Kataria (2014) makes three main points about a specific part of our paper (Maniadis, Tufano, and List 2014), namely about Tables 2 and 3. In our paper, we employ these tables in order to illustrate the idea that very inconclusive post-study probabilities that a tested phenomenon is true may result from novel, surprising findings. The main arguments in Kataria (2014) are the following:

First, if  $P(H_0)$  is unknown, as is often the case with economic applications, the post-study probability can lead to even worse inference than the Classical significance test, depending on the quality of the prior. Second, the simulation in Maniadis et al. (2014) ignores previous assessments of  $P(H_0)$  and instead utilizes a selective empirical setup that favors the use of post-study probabilities. ... [Third,] contrary to what Maniadis et al. (2014) argue, their results do not allow for drawing general recommendations about which approach is the most appropriate. (Kataria 2014, [abs.](#))

We believe that our work might have been misunderstood by Kataria. Moreover, it seems that some of his claims are not supported by relevant empirical evidence.

---

1. University of Southampton, Southampton SO17 1BJ, UK.  
2. University of Nottingham, Nottingham NG7 2RD, UK.  
3. University of Chicago, Chicago, IL 60637.

In Maniadis, Tufano, and List (2014), our basic aim is to draw on the general problem of the credibility crisis in disciplines other than economics (Ioannidis 2005; Bettis 2012; Jennions and Moller 2002), and to convey the disquieting news to economists by relying on insights and tools from the life sciences literature. While conveying the troubling news, we also emphasize the good news that usually it takes only a few independent replications to advance considerably the credibility of empirical exercises. We wish to understand how confident one should be in the published empirical findings in economics. Simply put, we are not discarding classical significance testing, just arguing that we should be interpreting it accurately. For an educated assessment of the empirical evidence we need to know not just whether tests were significant but also the value of key variables such as research priors and statistical power. Admittedly, these variables are not easy to estimate, and in economics it is often, even typically, the case that there is not much relevant evidence. But this is exactly our point: We wished to show that if we wish to assess how confident we are in our findings, evidence is lacking in critical dimensions. Given the recent evidence pointing to non-replicability in several life sciences (Ioannidis 2012), such lack of evidence may cause serious questions to be raised about economics as well (see Ioannidis and Doucouliagos 2013; Alexander 2013).

Whereas Kataria claims that “for economic hypotheses, the unconditional probability  $P(H_0)$  is hardly ever known” (Kataria 2014, 8), we suggest that the issue of such knowledge accumulation needs to be regarded as endogenous. If the investigator’s frame of analysis disregards the variable  $P(H_0)$ , there is no need to estimate it. Other disciplines have developed meta-analytic methods that can be fruitfully employed in economics for estimating the relevant variables (Cooper, Hedges, and Valentine 2009). Replication has a key role in these methods.

To encourage such a structured approach, we illustrated with Tables 2 and 3, using Bayesian language, the fact that we should be cautious of new evidence and—as we argue later in Maniadis, Tufano, and List (2014)—that we should also increase our efforts to replicate original studies. We clearly note in the paper that the combinations of parameter values used in Tables 2 and 3 should be thought of as applying to novel and surprising findings (Maniadis, Tufano, and List 2014, 278, 286 n. 27). So these combinations were *truly selected* to illustrate what happens in the case of such findings. Moreover, we acknowledged the difficulty of pinpointing those combinations exactly (*ibid.*, 286). Essentially, the degree to which our discipline is characterized by such combinations of priors and power is an empirical question. We hope that the message of the tables itself will encourage work on this underexplored question. Once more, we view as one of our key messages that we lack sufficient evidence to evaluate the credibility of much work in our field. We join others in prompting economists to grapple with such questions as: What is

a reasonable estimate for the typical prior in each subfield of economic research? What is the typical power of a research study? How common is replication in economics and how common should it be?

Given the scarcity of relevant empirical studies, we find the particular configurations suggested by Kataria (2014) somewhat unsupported by the evidence. In particular, there seems to be no empirical foundation for the claims that “effect sizes follow a standard normal distribution centered at zero and...scientists only detect and consider effect sizes  $|d| > 0.2$  as relevant” (Kataria 2014, 6). Despite this, Kataria claims that “the neighborhood of  $P(H_0) \approx 0.16$  ... is appreciated to be a more realistic estimate” (ibid., 7). Estimating  $P(H_0)$  is a difficult empirical question that would require much more research. With respect to power, Kataria mentions evidence from the related field of psychology, namely Joseph Rossi (1990), who estimated that the average power for medium effect sizes is equal to 0.57. However, it is not clear on which evidence the assumption of medium effect sizes is based. Furthermore, more recent evidence reveals that typical power in psychology is about 0.35, even if we assume that the average effect size  $|d|$  is equal to 0.5 (Bakker, van Dijk, and Wicherts 2012).

The spirit of our paper is to encourage work such as the very recent paper by Le Zhang and Andreas Ortman (2013). They retrospectively estimated the power of several experimental designs reported in Christoph Engel’s meta-analysis of dictator games (Engel 2011), and they found that the median level of power was less than 0.25. It is important to note the critical role of meta-analysis for generating this piece of new evidence. The point is not to argue in the absence of evidence but to try to accumulate the necessary evidence. As economists, we hope that our field is very credible, but we need to provide empirical evidence using the relevant tools.

At this point we need to acknowledge the important issue of “previous assessments of  $P(H_0)$ ,” although Kataria mentioned it without justification. As we said in Maniadis, Tufano, and List (2014), we aimed to make a claim about novel, surprising results. We do believe that many types of economic research are more grounded in theory than research in other social sciences, so for them “surprising” results may not be as important for publication. In fact, Brad DeLong and Kevin Lang (1992) found that  $P(H_0)$  is very close to zero for a set of hypotheses published in top economic journals in the 1980s. If their interpretation—that the referee process somehow manages to filter true associations—is correct, that would be reassuring for the credibility of the economics profession. As DeLong and Lang (1992) acknowledge, however, there are alternative interpretations for their findings, such as the existence of selection issues and data mining in the discipline, so their optimistic interpretation should be taken with caution. There is a need for further research on the matter, following the seminal analysis of DeLong and Lang

(1992). We are particularly interested in the field of experimental economics, where we worry that “surprising” findings might be more frequently published.<sup>4</sup>

From the previous arguments it should be clear that in Maniadis, Tufano, and List (2014) we did not put forward any general recommendation about which inference approach, Classical or Bayesian, is the most appropriate. In fact, in the context of the current “publish or perish” culture (see, e.g., Fanelli 2010) and the related structure and incentives of the economics knowledge system (Oswald 2007; Glaeser 2008; Young, Ioannidis, and Al-Ubaydli 2008), we merely resort to Bayesian language to argue in favor of a much more careful interpretation of Classical inference.

Summing up, we believe that studying systematically the factors that affect the credibility of empirical findings might have an important role to play in economics. Meta-analysis and Bayesian tools are of central importance for conceptualizing the problem and quantifying key variables, and should not be ignored by economists. Our point was not to argue in favor of a specific configuration of parameter values, but to show that we cannot ignore factors such as priors and power, because if we do, something can go very wrong with economic research.

## References

- Alexander, Ruth.** 2013. Reinhart, Rogoff... and Herndon: The Student Who Caught Out the Profs. *BBC News*, April 20. [Link](#)
- Bakker, Marjan, Annette van Dijk, and Jelte M. Wicherts.** 2012. The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science* 7(6): 543-554.
- Bettis, Richard A.** 2012. The Search for Asterisks: Compromised Statistical Tests and Flawed Theories. *Strategic Management Journal* 33(1): 108-113.
- Cooper, Harris, Larry V. Hedges, and Jeffrey C. Valentine,** eds. 2009. *The Handbook of Research Synthesis and Meta-Analysis*, 2nd ed. New York: Russell Sage Foundation.
- DeLong, J. Bradford, and Kevin Lang.** 1992. Are All Economic Hypotheses False? *Journal of Political Economy* 100(6): 1257-1272.
- Engel, Christoph.** 2011. Dictator Games: A Meta Study. *Experimental Economics* 14(4): 583-610.
- Fanelli, Daniele.** 2010. Do Pressures to Publish Increase Scientists’ Bias? An Empirical Support from US States Data. *PLoS ONE* 5(4): e10271. [Link](#)

---

4. We would welcome more empirical evidence on this and related issues.

- Glaeser, Edward L.** 2008. Researcher Incentives and Empirical Methods. In *The Foundations of Positive and Normative Economics: A Handbook*, eds. Andrew Caplin and Andrew Schotter, 300-319. New York: Oxford University Press.
- Ioannidis, John P. A.** 2005. Why Most Published Research Findings Are False. *PLoS Medicine* 2(8): e124. [Link](#)
- Ioannidis, John P. A.** 2012. Why Science Is Not Necessarily Self-Correcting. *Perspectives on Psychological Science* 7(6): 645-654.
- Ioannidis, John P. A., and Chris Doucouliagos.** 2013. What's to Know About the Credibility of Empirical Economics? *Journal of Economic Surveys* 27(5): 997-1004.
- Jennions, Michael D., and Anders P. Møller.** 2002. Relationships Fade with Time: A Meta-Analysis of Temporal Trends in Publication in Ecology and Evolution. *Proceedings of the Royal Society B* 269: 43-48.
- Kataria, Mitesh.** 2014. One Swallow Doesn't Make a Summer: A Comment on Zacharias Maniadis, Fabio Tufano, and John List. *Econ Journal Watch* 11(1): 4-10. [Link](#)
- Maniadis, Zacharias, Fabio Tufano, and John A. List.** 2014. One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *American Economic Review* 104(1): 277-290.
- Oswald, Andrew J.** 2007. An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica* 74(293): 21-31.
- Rossi, Joseph S.** 1990. Statistical Power of Psychological Research: What Have We Gained in 20 Years? *Journal of Consulting and Clinical Psychology* 58(5): 646-656.
- Young, Neal S., John P. A. Ioannidis, and Omar Al-Ubaydli.** 2008. Why Current Publication Practices May Distort Science. *PLoS Medicine* 5(10): e201. [Link](#)
- Zhang, Le, and Andreas Ortmann.** 2013. Exploring the Meaning of Significance in Experimental Economics. *University of New South Wales Australian School of Business Research Paper* 2013-32. [Link](#)

## About the Authors



**Zacharias Maniadis** is Assistant Professor of Economics at the Economics Division, School of Social Sciences, University of Southampton. His research interests include micro-economics, experimental economics, behavioral economics, and political economy. His email address is [Z.Maniadis@soton.ac.uk](mailto:Z.Maniadis@soton.ac.uk).



**Fabio Tufano** is Assistant Professor of Economics at the University of Nottingham. His research interests include psychology and economics, experimental economics, and game theory. His email address is [fabio.tufano@nottingham.ac.uk](mailto:fabio.tufano@nottingham.ac.uk).



**John A. List** is Homer J. Livingston Professor in Economics and the College at the University of Chicago. His research interests include experimental economics, field experiments, education, youth violence, economics of charity, environmental economics, experiments in firms, multi-unit auctions, and neuro-economics. His e-mail address is [jlist@uchicago.edu](mailto:jlist@uchicago.edu).

[Go to archive of Comments section](#)  
[Go to January 2014 issue](#)



Discuss this article at Journaltalk:  
<http://journaltalk.net/articles/5816>