Editorial

Towards a New Paradigm for Statistical Evidence in the Use of *p*-Value

Muhammad Ishaq Bhatti * and Jae H. Kim *

La Trobe Business School, La Trobe University, Melbourne, VIC 3086, Australia
* Correspondence: i.bhatti@latrobe.edu.au (M.I.B.); J.Kim@latrobe.edu.au (J.H.K.)

As the guest editors of this Special Issue, we feel proud and grateful to write the editorial note of this issue, which consists of **seven** high-quality research papers. We are incredibly grateful to the colleagues who submitted their papers and to the referees who provided thoughtful and constructive feedback within the tight deadlines provided by the journal. We acknowledge the hard work and commitments of the authors, who have implemented revisions that addressed all essential comments and suggestions of the referees. As guest editors, it was a pleasure for us to observe this prompt, collegial and constructive reviewing process. In our tasks and efforts, we were assisted by the efficient support provided by the Editor-in-Chief, Marc Paolella, and the assistant editor.

This Special Issue deals with problems of statistical inference and the use of *p*-values. Recently, the issue of the use of *p*-values in various scientific investigations and data analytics techniques has raised questions regarding the validity of statistical decision-making in social sciences including the business and economics disciplines. It is common practice among practitioners and researchers to make statistical decisions exclusively by using the "*p*-value < 0.05" criterion, regardless of sample size, statistical power and/or expected loss function underlying the selected models. Some of the well-known scholars have raised serious concerns about this practice and have warned that the use of "*p*-value" may lead to wrong decisions and give distorted scientific results. As an example, we quote a statement made by the American Statistical Association (Wasserstein and Lazar 2016) and the presidential address given by the American Finance Association (Harvey 2017). A few studies have commented on this issue by presenting empirical evidence, such as the paper by Keuzenkamp and Magnus (1995) and McCloskey and Ziliak (1996) for economics, Fazal et al. (2020) for energy, Kim et al. (2018) for accounting, and Kim and Choi (2017) for finance, among others.

The problem has become more challenging with increasing availability of large data sets. In particular, it is widely recognized that statistical significance (based on the conventional *p*-value criterion) is becoming irrelevant for big data (see, for example, Gandomi and Haider 2015). To this end, Rao and Lovric (2016) maintain that "the 21st century researchers work towards a '*paradigm shift*' in testing statistical hypothesis". There are calls that the researchers conduct more extensive exploratory data analysis before inferential statistics are considered for decision-making (see, for example, Leek and Peng 2015; Soyer and Hogarth 2012). There are even calls that the use of statistical significance based on the *p*-value criterion should be abandoned (Wasserstein et al. 2019). In light of these criticisms and calls for change, the Special Issue has been proposed.

The present Special Issue of *Econometrics* is a collection of seven excellent papers that address some of the following topics:

1. New or alternative methods of hypothesis testing such as estimation-based method (e.g., confidence interval), predictive inference, and equivalence testing.
2. Application of adaptive or optimal level of significance to business decisions.
3. Decision–theoretic approach to hypothesis testing and its applications.
4. Compromise between the classical and Bayesian methods of hypothesis testing.

5. Exploratory data analysis for large or massive data sets.
6. Critical review papers on the current practice of hypothesis testing and future directions in business.

This Special Issue begins with an inaugural article by Richard Startz entitled, **"Not p-Values, Said a Little Bit Differently"**, which is an important contribution toward the ongoing discussion about the use and/or misuse of *p*-values. Numerical examples are presented which demonstrate that a *p*-value can, as a practical matter, give you a different answer than the one that you want. Further contributions to the topic come from Thomas Dyckman and Stephen A. Zeff on **"Important Issues in Statistical Testing and Recommended Improvements in Accounting Research"**. This paper proposes improvements to both the quality and execution of research related to statistical inference in developing statistical tests which address the limitations in existing literature. They explore the situational effects of "data carpentry", alternatives to winsorizing, and suggest necessary improvements instead of relying on a study's calculated "*p*-values".

One of the many highlights of this Special Issue is the paper titled **"Interval-Based Hypothesis Testing and Its Applications to Economics and Finance"** authored by Jae Kim and Andrew P. Robinson. This paper tackles a long-standing literature review on interval-based hypothesis testing (such as tests for minimum-effect, equivalence, and non-inferiority) widely used in biostatistics, medical science, and psychology. It presents the methods in the contexts of a one-sample *t*-test and a test for linear restrictions in a regression. The paper employs testing for market efficiency, validity of asset-pricing models, and persistence of economic time series. Authors argue that, from the point of view of economics and finance, interval-based hypothesis testing provides more sensible inferential outcomes than those based on point-null hypothesis. It proposes interval-based tests which can be routinely used in empirical research in business, as an alternative to point null hypothesis testing, especially in the new era of big data.

Another paper addressing a similar issue is written by David Trafimow, entitled **"A Frequentist Alternative to Significance Testing, p-Values, and Confidence Intervals"**. In this article David begins his debate about null hypothesis significance testing, *p*-values without null hypothesis significance testing, and confidence intervals. The first major section addresses some of the main reasons these procedures are problematic and concludes that none of them are satisfactory. However, there is a new procedure, termed the a priori procedure (APP), which validly aids researchers in obtaining sample statistics that have acceptable probabilities of being close to their corresponding population parameters. The second major section provides a description and review of APP advances. Not only does the APP avoid the problems that plague other inferential statistical procedures, but it is easy to perform, too. Although the APP can be performed in conjunction with other procedures, the present recommendation is that it be used alone.

The fifth important paper is the contribution of Jan Magnus, who addresses the issue of the use of *t*-ratios. The title of this paper is **"On Using the t-Ratio as a Diagnostic"**, in which the author points out that *tests* and *diagnostics* are the two uses of *t*-ratios in econometrics. The paper proposes a new pretesting method *model averaging* over *t*-ratio and pretest estimators.

The sixth paper of the Special Issue is authored by John Quiggin with the title **"The Replication Crisis as Market Failure"**. Adopting a microeconomic approach, John's paper begins with the observation that the constrained maximization central to model estimation and hypothesis testing may be interpreted as a kind of profit maximization. The output of estimation is a model that maximizes some measure of model fit, subject to costs that may be interpreted as the shadow price of constraints imposed on the model. The replication crisis may be regarded as a market failure in which the price of "significant" results is lower than would be socially optimal.

The seventh paper is on the pedagogy of econometrics, entitled "**Teaching Graduate (and Undergraduate) Econometrics: Some Sensible Shifts to Improve Efficiency, Effectiveness, and Usefulness**" by Jeremy Arkes. According to Wasserstein et al. (2019),

"Statistics education will require major changes at all levels to move to a post '$p < 0.05$' world". As educators, we will need to rethink the way we train the future decision-makers, especially in the big data era where the p-value criterion is no longer relevant. Jeremy proposes a range of critical points on the issue of teaching econometrics, including the problem related to the p-value, maintaining that the teaching of graduate (and undergraduate) econometrics needs to be revamped.

We are very thankful to all authors, who have made considerable efforts to meet the standards of the journal. We believe that this Special Issue has been very successful in attracting seven high-quality contributions from established and well-known scholars in their prospective fields. The journal has provided an open access-publishing facility to the contributors, which is a realistic option for our discipline. We hope this Special Issue will stimulate further research and novelty in understanding a new paradigm for statistical evidence of p-value insights. We take this opportunity to thank the numerous reviewers who have greatly contributed to the quality of the published papers. Finally, we thank the editor-in-chief, Marc Paolella, and the team of assistant editors without whose devotion this Special Issue would not have been produced in such a smooth and well-managed manner.

Conflicts of Interest: The authors declare no conflict of interest.

References

Fazal, Rizwan, Syed Aziz Ur Rehman, Atiq Ur Rehman, Muhammad Ishaq Bhatti, and Anwar Hussain. 2020. Energy-Environment-Economy causal nexus in Pakistan: A Graph Theoretic Approach. *Energy* 214: 118934. [CrossRef]

Gandomi, Amir, and Murtaza Haider. 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* 35: 137–44. [CrossRef]

Harvey, Campbell R. 2017. Presidential Address: The Scientific Outlook in Financial Economics. *Journal of Finance* 72: 1399–440. [CrossRef]

Keuzenkamp, Hugo A., and Jan R. Magnus. 1995. On tests and significance in econometrics. *Journal of Econometrics* 67: 103–28. [CrossRef]

Kim, Jae H., and In Choi. 2017. Unit Roots in Economic and Financial Time Series: A Re-evaluation at the Decision-based Significance Levels. *Econometrics* 5: 41, This article belongs to the Special Issue *Celebrated Econometricians: Peter Phillips*. [CrossRef]

Kim, Jae H., Kamran Ahmed, and Philip Inyeob Ji. 2018. Significance Testing in Accounting Research: A Critical Evaluation based on Evidence. *Abacus* 54: 524–46. [CrossRef]

Leek, Jeffrey T., and Roger D. Peng. 2015. Statistics: *P* values are just the tip of the iceberg. *Nature* 7549: 520–612. [CrossRef] [PubMed]

McCloskey, Deirdre N., and Stephen T. Ziliak. 1996. The standard error of regressions. *Journal of Economic Literature* 34: 97–114.

Rao, Calyampudi Radhakrishna, and Miodrag M. Lovric. 2016. Testing Point Null Hypothesis of a Normal Mean and the Truth: 21st Century Perspective. *Journal of Modern Applied Statistical Methods* 15: 2–21. [CrossRef]

Soyer, Emre, and Robin M. Hogarth. 2012. The illusion of predictability: How regression statistics mislead experts. *International Journal of Forecasting* 28: 695–711. [CrossRef]

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. The ASA's statement on p-values: Context, process, and purpose. *The American Statistician* 70: 129–33. [CrossRef]

Wasserstein, Ronald L., Allen L. Schirm, and Nicole A. Lazar. 2019. Moving to a world beyond "$p < 0.05$". *The American Statistician* 73: 1–19.