*International Journal of Research and Review*
*Vol. 10; Issue: 7; July 2023*
*Website: www.ijrrjournal.com*
*Short Communication* *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

# Improving the Reproducibility in Forecasting Research

**P. Udhaya**

Guest Lecturer in Mathematics, Jawaharlal Nehru Rajkeeya Mahavidyalaya, Port Blair, Andaman & Nicobar Island

**DOI: https://doi.org/10.52403/ijrr.202307110**

## ABSTRACT

The value of reproduction is recognized in many scientific fields. Reproducibility is a necessary condition for reliability because the inability to reproduce results indicates that methods are not sufficiently specified, thus preventing replication. This article describes how two independent teams of researchers attempted to reproduce the empirical findings of an important study, "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy" (Miller & Williams, 2003, IJF). Teams of researchers proceeded systematically, reporting results before and after receiving clarifications. These inconsistencies have led to differences in conclusions about the conditions under which seasonal damping surpasses classical decay. The study addresses forecasting methods using a flow chart. It is argued that this approach to method documentation complements the provision of computer code by being accessible to a wider audience of forecasting practitioners and researchers. The importance of this research lies not only in its lessons for seasonal forecasting, but also in its approach to the reproduction of forecasting research.

***Keywords:*** Forecasting practice, Reproduction, Seasonal Forecasting, Empirical analysis

## INTRODUCTION

Replication of research is considered one of the basic criteria for determining its quality (Pulverer 2015; Van Bavel et al. 2016). The low percentage of studies in the various disciplines that meet the replication test in its various meanings as described below (Baker 2016; Ioannidis 2005) helps to explain the difficulty faced by the authors of articles in presenting synergistic models that incorporate cumulative research findings over time. The literature addresses several specific factors that explain the low replication rate and offers solutions. Our article attributes part of the problem to research procedures that have not been updated based on current information, and offers options for reducing the extent of the reproducibility problem through such updating.

In the absence of replication, scientific claims rest on the results of single, 'one shot', studies and hence carry risks and limitations. Researchers may have inadvertently made errors in theirapplication of methods. They may have made mistakes in data entry, committed arithmetic or data transcription errors or written computer code that contains bugs. They may also have made assumptions that are not stated explicitly and their findings may be sensitive to changes in these assumptions. Other assumptions, and even further errors, may be embedded in commercial software so that researchers are unaware of them (McCullough, 2000). In addition, results may apply only to the specific data that have been analysed and hence will be subject to sampling error. When statistically insignificant results are obtained, researchers may be tempted to "hunt for p-values less than 0.05" (Hubbard & Armstrong, 1994) andhence inflate the true probability of committing type I errors. This problem is avoided by replication studies, as

statistical significance is not a measure of replicability. Finally, the extent to which the findings generalize to situations or populations beyond those investigated in the original study will be unknown.

These potential risks and limitations suggest a range of approaches to replication. Definitions of replicability vary across disciplines, but a special case is reproducibility. If findings are reproducible, then independent researchers are able to obtain the same results as the original study using the same data and the same methods. Reproducibility is a first step towards replication and so, if it cannot be achieved, the generalizability of findings is likely to be in doubt of course, perfect reproduction of results may not be possible. For example, improvements in the algorithms embedded in software may lead to differences between the original numbers reported and those obtained using later versions of the software. However, approximate reproducibility, discussed later in this paper, may still be attainable. Findingsthat have been successfully reproduced have a much lower risk of being subject to human error. Further, the process of trying to reproduce findings is likely to reveal the extent to which the original results were based on unstated assumptions and hence the extent to which the findings will change if alternative assumptions are made.

Other easier alternatives have been proposed for measuring reproducibility, among them using triangulation to amass data (Munafo and Smith 2018), crowd sourced testing of a research hypothesis, including the cross-referencing of data from independent studies (Landy et al. 2020), exploratory and confirmatory factor analysis studies, obligatory detailed recording of the research process and full transparency for all those wishing to examine it (Field et al. 2020). The stringency demonstrated in these approaches increases measurement validity but at the same time may entail exposure to the novelty effect, if the data bases are composed solely of short-term measurements.

This study is about the process of reproducing results in forecasting research. We describe the process whereby two independent teams of researchers attempted to reproduce the findings of an award-winning study, "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy". We then identify issues that arose during the process and discuss how these issues may be resolved. The remainder of the article is organized as follows: in the next section, the relationship between reproducibility and replicability is discussed in more detail. In Section 3, the original research is described, the process of reproducing the results and the sources of discrepancies is explained and the impact of these differences on Miller & Williams' findings are discussed. A more detailed explanation of this process is given in table 1. Section 4 compares different approaches to the specification of forecasting methods and Section 5 concludes the study. A comprehensive flowchart of the forecasting process is given in figure 3.

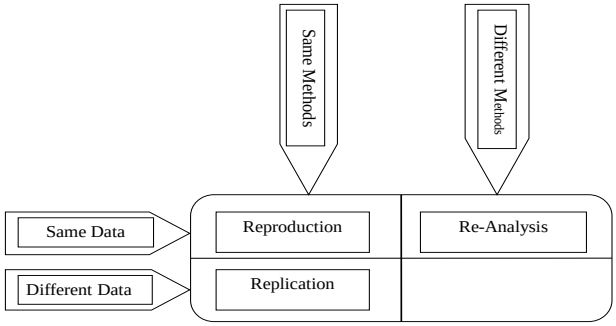
**Reproducibility and Replicability:**

Following on from the discussion in the previous section, we propose the following definitions of reproducibility and replicability in forecasting research. If results are reproducible then independent researchers are able to obtain the same numerical results by repeating the original study using the same methods on the same data. If findings are replicable then independent researchers are able to reach the same qualitative conclusions by repeating the original study using the same methods on different data. It should be possible for independent researchers to reproduce or replicate without any additional information from the author of the original study (King, 1995).

Evan schitzky and Armstrong (2010) use the term "re-analysis" to refer to an application of different methods on the same data or a sub-sample of the data. This constitutes a third category, in addition to "reproduction"

and "replication", as shown in Figure 1 below.

Figure: 1
Reproduction, Replication and Re-Analysis




Similar distinctions between reproducibility and replicability have been drawn in other scientific disciplines (e.g. in psychology by Asendorpf et al., 2013). However, it should be noted that these terms are sometimes used differently by other authors. For example, Drummond (2009) used the terms in the opposite way to the above definitions. Evan schitzkyet al. (2007), used the term "replication with extension" to indicate replication (in our terminology) but with a greater emphasis on generalisation.

Reproducibility is a necessary condition for replicability. An inability to reproduce the numerical results of a study implies that the methods used in that study have been insufficiently specified, thereby precluding replication. However, it is not a sufficient condition because the availability of further data meeting the necessary conditions is also required for a replication study to be conducted and for the qualitative findings to be replicated (e.g., in a forecasting context, method A is more accurate than method B under certain conditions).

Another important issue that has not been addressed in forecasting research is 'exact reproducibility'. Does precision to, say and the second decimal place only but not to the third, constitute a reproduction of a previous result or not? Such differences may arise from the use of different optimisation algorithms in different software packages. In this paper, a further distinction is drawn between 'exact reproducibility' and 'approximate reproducibility'. Exact reproducibility corresponds to our previous definition of reproducibility. On the other hand, if it is claimed that findings are approximately reproducible to a certain percentage, then independent researchers should be able to obtain results that differ by no more than that percentage by repeating the original study.

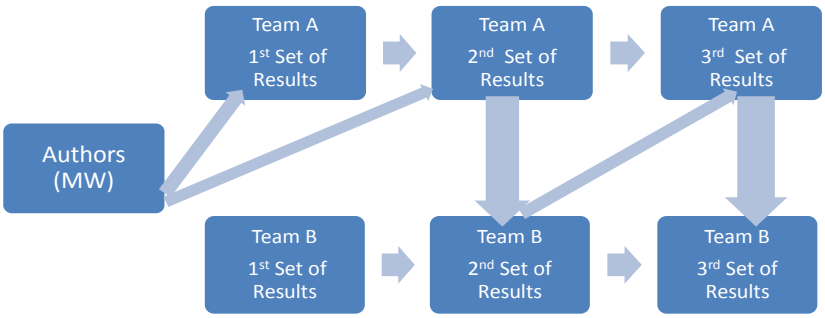
**The study by Miller & Williams:**

As previously discussed, the International Journal of Forecasting (IJF) is among those journals that support replication studies. Given that reproducibility is a necessary condition for replicability, we have focused on reproducing an important study published in the IJF, namely "Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy" (Miller & Williams, 2003). This article won an outstanding paper award, 2002- 2003, by the International Journal of Forecasting, and has been cited more than 25 times according to Google Scholar. It is also referred to in the well cited review by De Gooijer & Hyndman (2006) of the most important advancements in the recent history of forecasting.

The article by Miller & Williams (2003) is not untypical in its documentation of forecasting procedures. The authors give details of their dataset, methods for estimating seasonal factors and accuracy measures. They also provide some

information on parameter specification and prediction methods (although further details are needed on these topics, as discussed inTable 1).
The assumptions and methodological stages of the original research paper are explained in page 679 of Miller & Williams (2003). Both teams fully documented all the working methods and assumptions made in the process of generating the results. The reproduction process is depicted graphically in Figure 2 and is explained below.

Figure: 2
Reproduction Process




First, team A contacted Professors Miller and Williams (MW) seeking clarifications with regard to the data series. The authors provided team A with the exact 55 series out of the 66 monthly series used in the M1-competition which they used in their study. Subsequently, team A produced the first set of results by making various assumptions regarding those issues about which they were unclear Table 1. Then, they contacted MW again to resolvethe issues raised in the first run and, based on this new information, they produced the second set of results.

**Table: 1 Mean Absolute Percentage Errors for Team A and Team B**

| | | L-K recommended | | | J-S recommended | | |
|---|---|---|---|---|---|---|---|
| | **Horizon** | **CD** | **J-S** | **L-K** | **CD** | **J-S** | **L-K** |
| Team A 1st set of results | | 34 series | | | 8 series | | |
| | 1 | 8.99 | 9.05 | 8.91 | 8.22 | 7.56 | 7.82 |
| | 3 | 9.53 | 9.63 | 9.45 | 9.38 | 8.54 | 8.82 |
| | 6 | 10.20 | 10.40 | 10.13 | 10.82 | 9.86 | 10.16 |
| | 12 | 11.08 | 11.40 | 10.99 | 11.46 | 10.64 | 10.89 |
| | 18 | 11.04 | 11.20 | 10.74 | 14.00 | 14.23 | 13.95 |
| Team A 2nd set of results | | 30 series | | | 9 series | | |
| | 1 | 8.95 | 8.94 | 8.86 | 7.48 | 6.88 | 7.11 |
| | 3 | 9.49 | 9.52 | 9.45 | 8.58 | 7.82 | 8.06 |
| | 6 | 10.26 | 10.32 | 10.22 | 9.95 | 9.08 | 9.34 |
| | 12 | 11.05 | 11.00 | 11.03 | 10.85 | 10.09 | 10.33 |
| | 18 | 10.97 | 10.62 | 10.77 | 12.92 | 12.98 | 12.80 |
| Team A 3rd set of results | | 30 series | | | 9 series | | |
| | 1 | 7.15 | 7.24 | 6.73 | 7.53 | 8.98 | 8.93 |
| | 3 | 8.00 | 7.95 | 7.53 | 10.96 | 10.06 | 10.40 |
| | 6 | 9.49 | 9.51 | 9.18 | 11.48 | 11.41 | 11.34 |
| | 12 | 12.35 | 12.61 | 12.45 | 13.42 | 13.37 | 13.12 |
| | 18 | 13.96 | 14.10 | 14.05 | 14.92 | 15.08 | 14.65 |
| Team B 1st set of results | | 36 series | | | 8 series | | |
| | 1 | 7.21 | 7.32 | 6.90 | 8.22 | 9.38 | 9.55 |
| | 3 | 7.82 | 7.88 | 7.43 | 12.11 | 10.84 | 11.43 |
| | 6 | 9.21 | 9.25 | 8.86 | 12.66 | 12.53 | 12.59 |
| | 12 | 11.80 | 11.66 | 11.47 | 14.31 | 13.78 | 13.92 |
| | 18 | 13.34 | 13.06 | 12.90 | 15.39 | 15.19 | 15.17 |
| Team B 2nd set of results | | 30 series | | | 9 series | | |
| | 1 | 7.14 | 7.24 | 6.72 | 7.53 | 8.98 | 8.93 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | 3 | 7.99 | 7.95 | 7.52 | 10.96 | 10.06 | 10.40 |
| | 6 | 9.46 | 9.51 | 9.15 | 11.49 | 11.41 | 11.34 |
| | 12 | 12.31 | 12.61 | 12.42 | 13.43 | 13.37 | 13.12 |
| | 18 | 13.91 | 14.10 | 14.00 | 14.93 | 15.08 | 14.65 |
| Team B 3rd set of results | | 30 series | | | 9 series | | |
| | 1 | 7.14 | 7.24 | 6.72 | 7.53 | 8.98 | 8.93 |
| | 3 | 7.99 | 7.95 | 7.52 | 10.96 | 10.06 | 10.40 |
| | 6 | 9.46 | 9.51 | 9.15 | 11.49 | 11.41 | 11.34 |
| | 12 | 12.31 | 12.61 | 12.42 | 13.43 | 13.37 | 13.12 |
| | 18 | 13.91 | 14.10 | 14.00 | 14.93 | 15.08 | 14.65 |

| | | **CD or J-S recommended** | | | **All 55 series** | | |
|---|---|---|---|---|---|---|---|
| | **Horizon** | **CD** | **J-S** | **L-K** | **CD** | **J-S** | **L-K** |
| Team A 1st set of results | | 13 series | | | 55 series | | |
| | 1 | 5.37 | 5.32 | 5.39 | 8.024 | 7.954 | 7.919 |
| | 3 | 5.55 | 5.42 | 5.50 | 8.566 | 8.476 | 8.425 |
| | 6 | 5.94 | 5.82 | 5.92 | 9.284 | 9.236 | 9.142 |
| | 12 | 6.77 | 6.63 | 6.75 | 10.113 | 10.160 | 9.975 |
| | 18 | 7.86 | 7.66 | 7.80 | 10.715 | 10.804 | 10.511 |
| Team A 2nd set of results | | 16 series | | | 55 series | | |
| | 1 | 6.76 | 6.69 | 6.67 | 8.071 | 7.949 | 7.936 |
| | 3 | 6.99 | 6.92 | 6.86 | 8.615 | 8.486 | 8.470 |
| | 6 | 7.60 | 7.54 | 7.49 | 9.434 | 9.308 | 9.284 |
| | 12 | 8.61 | 8.49 | 8.51 | 10.306 | 10.122 | 10.181 |
| | 18 | 9.62 | 9.64 | 9.44 | 10.896 | 10.725 | 10.717 |
| Team A 3rd set of results | | 16 series | | | 55 series | | |
| | 1 | 6.47 | 6.97 | 6.17 | 7.016 | 7.446 | 6.929 |
| | 3 | 7.34 | 7.49 | 7.04 | 8.291 | 8.160 | 7.860 |
| | 6 | 7.45 | 7.42 | 7.25 | 9.222 | 9.211 | 8.969 |
| | 12 | 8.66 | 8.48 | 8.44 | 11.451 | 11.533 | 11.396 |
| | 18 | 10.14 | 10.01 | 9.93 | 13.004 | 13.071 | 12.950 |
| Team B 1st set of results | | 11 series | | | 55 series | | |
| | 1 | 5.86 | 5.74 | 5.86 | 7.085 | 7.301 | 7.078 |
| | 3 | 6.85 | 6.64 | 6.95 | 8.249 | 8.065 | 7.915 |
| | 6 | 7.01 | 6.88 | 7.08 | 9.270 | 9.251 | 9.045 |
| | 12 | 8.34 | 8.20 | 8.43 | 11.475 | 11.276 | 11.222 |
| | 18 | 9.83 | 9.71 | 9.92 | 12.939 | 12.698 | 12.633 |
| Team B 2nd set of results | | 16 series | | | 55 series | | |
| | 1 | 6.31 | 6.97 | 6.01 | 6.964 | 7.446 | 6.874 |
| | 3 | 7.18 | 7.49 | 6.88 | 8.239 | 8.160 | 7.804 |
| | 6 | 7.30 | 7.42 | 7.09 | 9.164 | 9.211 | 8.908 |
| | 12 | 8.50 | 8.48 | 8.29 | 11.388 | 11.533 | 11.330 |
| | 18 | 9.99 | 10.01 | 9.79 | 12.937 | 13.071 | 12.880 |
| Team B 3rd set of results | | 16 series | | | 55 series | | |
| | 1 | 6.47 | 6.97 | 6.17 | 7.010 | 7.446 | 6.922 |
| | 3 | 7.34 | 7.49 | 7.04 | 8.284 | 8.160 | 7.851 |
| | 6 | 7.45 | 7.42 | 7.25 | 9.209 | 9.211 | 8.955 |
| | 12 | 8.65 | 8.48 | 8.44 | 11.431 | 11.533 | 11.375 |
| | 18 | 10.13 | 10.01 | 9.93 | 12.978 | 13.071 | 12.923 |

On the other hand, team B generated their first set of results using only the information given in the original paper and the 111 series from the M1 dataset without any contact with MW. They selected the same series used by Team A, showing that MW had provided sufficient information to allow specification of the exact 55 series. Then, team A provided team B with the additional information gained through their communication with MW and, based on that, team B produced their second set of results.

After the first set of results were presented by team B and upon a review of the documented stages of their replication, team A found other discrepancies. In their third run, team A attempted to repeat what team B did, by amending their experimental structure to match the assumptions and methods of team B. These stages along with the results producedin each of them are explained in detail in Table 1. The notations used are the same as in Miller & Williams (2003).

As mentioned in Table 1 even after further communications among the two teams, there were still discrepancies between their results (team A third set of results and team B second set of results in Table 1). In

order to investigate this issue, each of the 55 series has been checked individually (manually) to identify the series for which the Mean Absolute Percentage Error (MAPE) results produced by the two teams were different.

**Specification of forecasting methods:**

It is common for authors of forecasting papers to include statements of methods, including assumptions, in words (textual descriptions). However, it may be very difficult for others to translate these words into an unambiguous form for reproduction of results, replication of findings or adaptation of methods. To address this issue, some alternative approaches are discussed in this section.

One way of presenting methods is through the use of flow charts. A flow chart is a type of diagram that presents a method in algorithmic form, showing the steps as boxes of various kinds, and their order by connecting them with arrows. They are used extensively in simulation modelling (e.g. Hayes, Leal, Gray, Holman, & Clarke, 2013) but not so widely in forecasting. Another alternative is that the code itself may be offered alongside an academic paper. The internet is a significant aid to those who wish to make their data and algorithms available, for example by the use of journals' electronic companions.

Both flowcharts and code have advantages and disadvantages in facilitating reproduction and replication. Making code available guarantees exact reproduction of results while a flowchart may allow for only approximate reproduction. Nevertheless, exact reproduction using provided code may conceal errors, which might otherwise be revealed if new independent code is developed. In replication some small changes to code may be needed to cater for new data sets (e.g. different sample sizes), but the effort involved in carrying out the replication will be relatively small. Providing a flowchart will necessitate the development of new code with the attendant dangers of introducing programming errors, which may be less easy to identify than in reproduction, given the absence of a set of earlier results based on the same data.

If the methods need to be adapted, using flowcharts and developing new code may be easier than adjusting code developed by other researchers. A flowchart is more accessible than code and requires only a basic understanding of the flowcharting rules and conventions. It is easy and quick to read and apprehend. On the other hand, using code requires an understanding of the language of the code, which may need a significant time to acquire. Another concern about provision of code is that people's knowledge will affect its accessibility. For example, there are fewer people today who are able to read code in APL than 30 years ago.

Flow charting and provision of code are not mutually exclusive. On the contrary, they are complementary. Some researchers may wish to reproduce or replicate without adaptation of methods. Other researchers may wish to experiment with adaptations of forecasting methods. Provision of flowcharts and code caters for both research audiences.

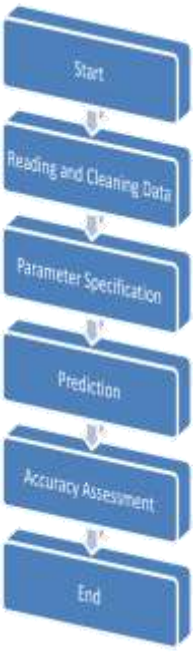
To summarise, textual description of methods and assumptions has been a common approach in forecasting studies. This approach was also adopted by MW in the research analysed in this paper. However, our results in Section 3 show that reproduction of the results of MW's research was not possible based on the information provided in the paper. As discussed in this sub-section, alternative approaches, such as flowcharts and provision of code, may facilitate reproduction, replication and adaptation. The application of flowcharting to the research presented in this paper will be discussed in the next sub-section.

**Flowchart for reproducibility:**

As explained, flowcharts are very accessible and easy to understand and, although they have not been widely used in forecasting studies, they can be easily implemented.

We have presented the detailed flowchart for the methods analysed in this article in Figure 3 using the information gathered from the authors of the original paper and the communications between the two teams. The flowchart consists of four blocks as shown below (Figure 3).

Figure: 3
Flow chart of the forecasting Process




The blocks could be used for a variety of forecasting approaches. For example, parameter specification for smoothing methods (which has been used here) includes initialisation and optimisation whereas, for the Box-Jenkins approach, it contains identification and estimation. Distinguishing these blocks in the code would also increase the clarity of the code and facilitate understanding and adaptation for reproducing the results and or replicating the findings.

## CONCLUSIONS AND IMPLICATIONS

In this study we have attempted to reproduce the results provided by Miller & Williams (MW, 2003). Our aim was to assess the feasibility and accuracy of doing so. It is important to emphasize that the methods in the MW article were not untypical in their fullness of documentation, compared to other papers in the forecasting literature. Hence, the MW article may be regarded as representative of method documentation in forecasting research.

We have worked in two teams (each of which attempted independently to reproduce the MW results) and in a structured way that allowed for the progressive accumulation of information relevant to the data and methods used in the MW study. Although the two teams reached almost the same results, those were different from the results provided by MW and we have not arrived at the same conclusions as the original article. This provides an example of where lack of reproduction of results matters in terms of replication of the findings and conclusions.

It is also important to note that the two teams did not achieve exact reproduction of each other's results, because of differences in software optimisation methods. Based on the outcomes of this work, we believe that there is considerable

scope for improving the reproducibility of forecasting research journal in general and articles published by the International Journal of Forecasting (IJF) in particular. The IJF requires that "for empirical studies, the description of the method and the data should be sufficient to allow for replication". However, in practice, it is uncommon for the reviewers or the editorial office to request details that are sufficient to reproduce the results. Consequently, there is an overreliance of the academic community on the goodwill of the authors of the original studies to answer simulation related queries, provide empirical data and clarify methodological issues.

In an attempt to enable other researchers to reproduce, replicate or adapt the methods used by MW, we have provided a fully documented flowchart of the methods in the article. We argued that flow-charts are accessible to a broader audience of forecasting practitioners and researchers than provision of code. However, we suggested that flow charts and codes are complementary in providing high level understanding and granular appreciation of forecasting methods. To that end, we have supplemented our paper with electronic companions that include both the flowcharts and the code written by the two teams of researchers.

We would like to close our paper by inviting other researchers to attempt to reproduce our results. This would enable the approach to reproducibility proposed in this study to be tested and commented upon by others. We also acknowledge that the issues discussed in this paper arise from a single research study and we would encourage researchers to attempt to reproduce other important forecasting studies and expand on the recommendations made in this study. Finally, and most importantly, we would encourage authors (including ourselves) to consider the issues of reproducibility and replication when documenting forecasting procedures and experimental structures employed for their research.

***Declaration by Authors***

**Acknowledgement:** None

**Source of Funding:** None

**Conflict of Interest:** The authors declare no conflict of interest.

## REFERENCES

1. Asendorpf, J. B., Conner, M., Fruyt, F. D. E., Houwer, J. A. N. D. E., Denissen, J. J. A., Fiedler, K., Vanaken, M. A. G. (2013), "Recommendations for Increasing Replicability in Psychology", European Journal of Personality, 27(2), 108–119.
2. Baker, M. (2016), "1,500 scientists lift the lid on reproducibility", Nature, 533(7604).
3. De Gooijer, J. G., & Hyndman, R. J. (2006), " 25 Years of Time Series Forecasting", International Journal of Forecasting, 22(3), 443–473.
4. Drummond, C. (2009), "Replicability is not Reproducibility: Nor is it Good Science. Proceedings of the Evaluation Methods for Machine Learning Workshop at the 26th ICML",Montreal, Canada.
5. Evanschitzky, H., & Armstrong, J. S. (2010), " Replications of forecasting research", International Journal of Forecasting, 26(1), 4–8.
6. Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J. S. (2007), "Replication Research's Disturbing Trend", Journal of Business Research, 60(4), 411–415.
7. Field, S.M., Wagenmakers, E.J., Kiers, H.A., Hoekstra, R., Ernst, A.F., van Ravenzwaaij, D. (2020), "The effect of preregistration on trust in empirical research findings: results of a registered report", Royal Society Open Science 7(4), 181351.
8. Hayes, A. J., Leal, J., Gray, A. M., Holman, R. R., & Clarke, P. M. (2013), " UKPDS Outcomes Model 2: A New Version of a Model to Simulate Lifetime Health Outcomes of Patients with Type 2 Diabetes Mellitus using Data from the 30-year United Kingdom Prospective Diabetes Study: UKPDS 82", Diabetologia, 56(9), 1925–33.
9. Hubbard, R., & Armstrong, J. S. (1994), "Replications and Extensions in Marketing – Rarely Published But Quite Contrary", International Journal of Research in Marketing, 11(3), 233–248.
10. Ioannidis, J. P. (2005), "Why most published research findings are false", PLoS medicine, 2(8), e124.

11. King G. (1995), "Replication, Replication", Political Science & Politics, 28(3), 443-452.
12. Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., (2020), "Crowd sourcing hypothesis tests: Making transparent how design choices shape research results", Psychological bulletin, 146(5), 451-479.
13. McCullough, B. (2000), "Is it Safe to Assume that Software is Accurate?", International Journal of Forecasting, 16(3), 349–357.
14. Miller, D. M., & Williams, D. (2003), "Shrinkage Estimators of Time Series Seasonal Factors and their Effect on Forecasting Accuracy", International Journal of Forecasting, 19(4), 669–684.
15. Munafo, M. R., & Davey Smith, G. (2018), "Robust research needs many lines of evidence", Nature, 553(7689), 399-401.
16. Pulverer, B. (2015), "Reproducibility blues", The EMBO journal, 34(22), 2721-2724.
17. Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016), "Contextual sensitivity in scientific reproducibility", Proceedings of the National Academy of Sciences, 113(23), 6454-6459.

How to cite this article: P. Udhaya. Improving the reproducibility in forecasting research. *International Journal of Research and Review*. 2023; 10(7): 950-958.
DOI: *https://doi.org/10.52403/ijrr.202307110*

******