

Referee report on entitled “Replicating ’Predicting the Present with Google Trends” by Hyunyoung Choi and Hal Varian, *The Economic Record*, Vol. 88, Special Issue, June, 2012, 2-9.”, Manuscript Number: MS-2350

## 1 Summary

This paper sets out a replication plan for the well-known paper by Choi and Varian (2012) that discusses the usefulness of a few applications of Google search data to forecast economic indicators in the short-term.

According to the Editors instructions to contributors, the replication plan should provide: (i) a general discussion of principles about how one should do a replication, (ii) an explanation of why the candidate paper was selected for replication, (iii) a replication plan that applies these principles to the candidate article, and (iv) a discussion of how to interpret the results of the replication (e.g., how does one know when the replication study replicates the original study).

While the author lays out the replication plan and does a satisfactory job of addressing (i)-(iv), I mention some broader aspects in Major comments that can enrich the discussion on how to do a replication in the context of using search data to forecast economic indicators. Also, the paper is hard to read due to the numerous typos, grammatical errors, and convoluted sentences. These are given in Minor comments.

## 2 Major Comments

While the replication plan did not require the author to replicate the study, I credit the author in trying to run the original code with the original data set on one of the time series, US car sales (as reported in Table 1) as the first ways to approach a replication.

1. A second way to do a replication according to the author would be to construct the dataset from scratch and write code based on the data used in Choi and Varian (2012). This is mostly dicussed in p2. para 1. I would like the author to discuss how recent studies (post Choi and Varian (2012)) have used Google trends for short-term forecasting. Is it useful to work with longitudinal serach data or ue higher frequency search data in short-term forecatsing? While these may reveal different results, richer datasets usually add value to forecasting exercises.

2. Another way to do a replication would be to use a similar/related search/query term for forecasting a particular economic indicator. An example of this is that Choi and Varian (2012) use searches on unemployment to better forecast unemployment claims whereas Baker and Fradkin (2011) have examined how job search relates to extensions of unemployment payments. How would one interpret results if a different query term is used for forecasting? I would imagine that not all similar/related search query terms will contribute to the predictive analysis in the same manner.
3. In all the aforementioned cases, it is important to mention that sub-sample analysis may reveal different results. A discussion of how turning points may affect forecasting accuracy may be useful here as discussed in Choi and Varian (2012)
4. I would like the author to discuss the other examples of economic indicators in Choi and Varian (2012) wherever necessary.

### 3 Minor comments

There are spelling mistakes and grammatical mistakes throughout the paper which are mentioned below. The author(s) should revise the paper thoroughly before submitting again.

1. p.1, para 1, line 1: delete 'one could try'
2. p.1, para 1, line 2: replace "includes" with 'include'
3. p.1, para 1, line 5: insert commas - after "show" and "example"
4. p.1, para 1, line 6: insert 'a' before "series"; replace "reflect" with 'reflects'
5. p.1, para 2: delete "with Google as affiliation"
6. replace "per cent" with '%' or 'percent' throughout the paper
7. Put a sentence somewhere referencing Table 1 rather than putting a placeholder in the text
8. p.1, para 4: unclear phrase "with typos explaining deviations"
9. p.2, para 1, lines 2-7: convoluted and long; insert 'a' before "replication"
10. p.2, para 1, line 11: It should read 'To get..'

11. p.2, para 1, line 12-13: You probably mean that exact replications cannot be made, however, results are similar to the original paper.
12. p.2, para 1, last 2 lines: grammatically incorrect
13. p.2, para 2, line 1: 'these kinds' instead of "this kind"
14. p.2, para 2, line 2: delete "happen"; 'provide' instead of "provides"; 'incentive for researchers' not "to"
15. p.2, para 2, line 3: delete "double and triple"
16. p.2, para 1, last line should be rewritten: you can suggest that Google trends has value added for short term predictability. It's long term significance remains to be seen as more data becomes available.
17. p.2, para 3 should be checked for grammatical errors
18. p.3, para 3, line 5: delete variables (it occurs twice)
19. p.3, para 3, last 3-4 lines: It is your interpretation that someone else reading the Choi and Varian (2012) will conclude that google search data adds predictive power at all times. How do you know this to be true? These kinds of sentences are speculative and should be deleted.
20. Rewrite p.3, last para, lines 1-3: should be revised accordingly.
21. As a general rule, footnotes should go after the period at the end of the sentence.