

Coupé, Tom

**Working Paper**

Replicating "Predicting the present with Google trends"  
by Hyunyoung Choi and Hal Varian (The Economic Record,  
2012)

Economics Discussion Papers, No. 2017-76

**Provided in Cooperation with:**

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

*Suggested Citation:* Coupé, Tom (2017) : Replicating "Predicting the present with Google trends" by Hyunyoung Choi and Hal Varian (The Economic Record, 2012), Economics Discussion Papers, No. 2017-76, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<https://hdl.handle.net/10419/169135>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>

**Discussion Paper**

No. 2017-76 | September 28, 2017 | <http://www.economics-ejournal.org/economics/discussionpapers/2017-76>

## Replicating “Predicting the present with Google trends” by Hyunyoung Choi and Hal Varian (The Economic Record, 2012)

*Tom Coupé*

**Abstract**

In this note, the author describes different ways one could try to replicate Choi and Varian (Predicting the present with Google trends, *The Economic Record*, 2012).

(Published in Special Issue [The practice of replication](#))

**JEL** A1 C1 C8 C53

**Keywords** Replication

**Authors**

*Tom Coupé*, ✉ University of Canterbury, Christchurch, New Zealand,  
[tom.coupe@canterbury.ac.nz](mailto:tom.coupe@canterbury.ac.nz)

**Citation** Tom Coupé (2017). Replicating “Predicting the present with Google trends” by Hyunyoung Choi and Hal Varian (The Economic Record, 2012). Economics Discussion Papers, No 2017-76, Kiel Institute for the World Economy. <http://www.economics-ejournal.org/economics/discussionpapers/2017-76>

In this note, I describe different ways one can replicate a paper and illustrate these ways by applying them to Choi and Varian (2012).

A replication can consist of 3 steps.

Step one is checking whether one can replicate the results of the paper using the data and code provided by the authors. One would expect that with the help of the authors one can replicate the various estimates reported in a paper. Such exact replication could be considered as successful if the vast majority of coefficients in a paper can be replicated exactly, with typos explaining deviations.

Second, one can try to collect the data and write the code based on the description in the paper one attempts to replicate. Ideally such replication would lead to the exactly the same numerical estimates though given the likely presence of implicit assumptions, small unreported data manipulations, updated data sources or misunderstandings by the person doing the replication, a replication could be deemed successful even if numerical estimates are not exactly replicated but qualitative conclusions (i.e. statistical and economic significance) would be confirmed.

The two abovementioned approaches to replication are important as knowing this kind of checks can happen provides an incentive to researchers to spend more efforts on avoiding mistakes, to double and triple check their own results and ideally, to make their code, data and details on how these were created, publicly available. At the same time, these two approaches to replication are rather narrow: they only check whether one get similar results if one does exactly the same thing for the same time period and the same country and the same data source. However, since most people will only read abstracts, introductions and/or conclusions, and since we all have the tendency (and wish) to extrapolate results of studies of specific situations to general laws of economics, 'broader' replication is also needed.

These more comprehensive checks would be replication attempts of the overall conclusions of paper one wished to replicate rather than of the exact numerical estimates reported in the paper. That is, does one find that for many countries, or series, or time periods, the general conclusion still holds. Such replication attempts are important as they provide an incentive for researchers to avoid cherry picking results and to do various robustness checks. In addition, such checks should push researchers to be very careful in their conclusions and make it clear to the reader that their results are for a specific dataset and setting rather than should be seen as an illustration of an economic law that is true always and everywhere. Such comprehensive replication efforts are similar in nature to what meta-analyses or literature surveys are doing.

Let's now turn to the paper we will use to illustrate the above steps, the paper by Choi and Varian (2012)

Choi and Varian (2012) includes 4 examples of macroeconomic statistics that can be forecasted more accurately if time series, reflecting the search intensity of terms related to the macroeconomic statistics, are included in the regression used to forecast the macroeconomic statistics. They show for example that if one augments an autoregressive model of U.S car sales, with series that reflect the evolution of search intensity for 'Trucks and SUVs' and

“Auto Insurance”<sup>1</sup>, both in-sample and out-of-sample forecast accuracy improves by about 10%. Other examples focus on the forecasting of US unemployment benefits claims, visitor arrivals to Hong Kong and Consumer confidence in Australia.

Choi and Varian (2012) has 1189 citations in Google Scholar. This is substantially more than Ettredge et al.(2005), the paper identified by Choi and Varian (2012) as the first to suggest using web search data in forecasting economic statistics, and substantially more than the early survey of this literature by Goel et al. (2010), also cited by Choi and Varian (2012). Ettredge et al.(2005) has 141 citations in Google Scholar while Goel et al. (2010) has 385 citations in Google Scholar. One possible explanation for the relative popularity of this paper is the fact that it has been written by two authors with Google as affiliation, and hence could be seen as some kind of ‘official’ paper connecting economic forecasting and web search intensity. An alternative explanation is that the Choi and Varian paper (2012) is more optimistic about how useful search intensity can be for prediction. Choi and Varian (2012) conclude: ‘We have found that simple seasonal AR models that include relevant Google Trends variables tend to outperform models that exclude these predictors by 5 per cent to 20 per cent.’ While Goel et al. (2010)’s conclude: ‘We conclude that in the absence of other data sources, or where small improvements in predictive performance are material, search queries provide a useful guide to the near future’.

The Choi and Varian (2012) paper can be replicated in the various ways described above, each way having a different goal in mind. Table 1 compares the results reported in the paper with various alternative approaches, focusing on the example of forecasting US car sales.

Table 1 about here

Column II checks whether one can replicate the results of the paper using the data and code provided on Varian’s website<sup>2</sup>. In this case, using Choi and Varian (2012)’s data and code for the US car sales I get exactly the same results. If a further check of the other examples in the paper would also result in this conclusion, the first step of our replication would be deemed successful.

In column III of table I, I make an attempt to collect the data and write the code based on the description in the Choi and Varian’s paper, rather than using the code and data they provide on Varian’s website. For papers based on Google’s search intensity, this kind of replication is unlikely to lead to exactly the same numerical estimates. As Choi and Varian (2012) note in their paper: “Note that Google Trends data is computed using a sampling method, and the results therefore vary a few per cent from day to day.” Hence, it’s unlikely that 6 or so years after Choi and Varian created their search intensity series, I get exactly the same series as they got. For the same reason, this replication will not be replicable as on different days I get slightly different numbers when downloading the search intensity series. Similarly, the US Sales series data available from the link provided in the paper are slightly different from the data provided on Varian’s website<sup>3</sup>. Column III shows that when I use the data I downloaded

---

<sup>1</sup> These are categories of searches rather than specific search terms, hence this is the evolution of the intensity for search terms that fall into the category ‘Trucks and SUVs’ and “Auto Insurance”.

<sup>2</sup> <http://www.sims.berkeley.edu/~hal/Papers/2011/Data.zip>

<sup>3</sup> The link to the data provided in footnote 2 of Choi and Varian (2012) refers to <https://www.census.gov/retail/marts/www/timeseries.html>

from Google Trends I get slightly different estimates but the coefficients are sufficiently similar, both in the statistical and economic sense, suggesting that the Choi and Varian (2012) paper would also pass this second replication step.

Somebody reading only the conclusion of Choi and Varian (2012) that “We have found that simple seasonal AR models that include relevant Google Trends variables tend to outperform models that exclude these predictors by 5 per cent to 20 per cent” might well be convinced that Google Search helps to predict everything and everywhere. Hence, the third step in our replication plan tries to check whether Choi and Varian (2012)’s conclusion survives when applied to other settings than those checked in the paper.

In column IV, I extend the dataset from 2011 to 2017, and find that the predictive impact of adding search intensity terms gets weaker. While the coefficients of the search intensity variables are significant, the adjusted  $R^2$  increases by less than 5%. Similarly, using Choi and Varian’s model and time period on NZ car sales data shows insignificant coefficients for both search intensity variables, though the increase in adjusted  $R^2$  is still more than 5% (column V). Extending the data up to 2017 for NZ further shows only a negligible increase in adjusted  $R^2$  (column VI). Of course these are just some counterexamples<sup>4</sup>. Varying the countries or time periods further, one might find that differences due to the longer period or the NZ data are the exception rather than the rule. Hence, a replication should check systematically, varying one aspect of the original study, whether adding Google search intensity series adds predictive power.

For example, one could start by collecting data for a given series for many countries and estimate for all countries the same model suggested by Choi and Varian (2012), i.e. add search intensity data for ‘Trucks and Suv’s and for ‘Auto Insurance’ to an AR(1,12) model<sup>5</sup>. This would allow to investigate to what extent there is heterogeneity in the contribution of these two search intensity series. And if such heterogeneity is found, one could try to explain this heterogeneity. For example, in countries where there is a greater number of searches (like the US) the predictive contribution of search intensity series could be bigger than in countries with a smaller number of searches (like New Zealand).

Similarly, one could vary other aspects of the Choi-Varian (2012) paper. One could continue to focus just on the US but check how changing the model affects the contribution of the search intensity series. One could for example, include additional autoregressive terms or lags

---

which gives the “Adjusted Monthly Sales for Retail and Food Services” and the adjustment coefficients which are rounded at 3 digits. Multiplying these two quantities gives what should be the unadjusted series used by Choi and Varian (2012). The unadjusted time series of the estimates can also be obtained from the Census bureau

[https://www.census.gov/econ/currentdata/dbsearch?program=MARTS&startYear=1992&endYear=2017&categories%5B%5D=441&dataType=SM&geoLevel=US&notAdjusted=1&submit=GET+DATA&releaseScheduleId=\]](https://www.census.gov/econ/currentdata/dbsearch?program=MARTS&startYear=1992&endYear=2017&categories%5B%5D=441&dataType=SM&geoLevel=US&notAdjusted=1&submit=GET+DATA&releaseScheduleId=])

These two ‘unadjusted’ series are not exactly the same but very similar, differences could be related to rounding. Both series are slightly different from the data provided on Varian’s website. The search intensity data on Varian’s website are also standardized in a different way than the one’s available from Google Trends.

<sup>4</sup> These counterexamples are only based on in-sample forecasting performance. Choi and Varian (2012) also check out-of-sample forecast performance.

<sup>5</sup> One example of a step in this direction is [Tuhkuri \(2016\)](#) who studies not only how Google Search can help predict unemployment at the US level but also at the level of the various different US states.

of other series. Or one could vary the time period of the analysis, applying the Choi-Varian (2012) model to shorter or longer periods of data, or to non-overlapping various sub-periods.

It is hard to predict what the outcome of this third replication step would be though my guess is that Choi and Varian (2012)'s results would be true not just for the examples they tested, but rather would be true for some places and some time periods, rather than always and everywhere. Choi and Varian (2012)'s conclusion is written such that it allows for some uncertainty: 'relevant Google Trends variables **tend [bold added]** to outperform models that exclude these predictors by 5 per cent to 20 per cent'<sup>6</sup>. So the result of the third step replication would help to quantify the '**tend [bold added]** to outperform' from Choi and Varian's conclusion. That is, once the study's internal replication is successful (step 1 and 2), we know that under some circumstances, adding search intensity series improves forecasts. If step 3 would come to the conclusion that adding search intensity series improves forecasts only for the cases they describe in the paper, one could argue that step 3 of the replication was not successful, that their results are not externally replicable. However, just like it is unlikely that a single study will consider all possible circumstances, it is unlikely that any replication would be able to consider all possible circumstances. Hence, rather than leading to a binary conclusion (the external replication is successful or not successful), the replication would estimate how often, and in what circumstances, adding search intensity series improves forecasts.

So far, I have focused on a three steps replication plan, where the first two steps focus on the internal validity of the paper, while the third step focuses on the external validity of the paper. These three steps are all focusing on 'academic' replication, however. But replication could also leave the ivory tower and check whether a paper is replicable in 'real life', that is, are the conclusions of a paper, after the paper has been made public, actually used by decision makers? In the case of Choi and Varian (2012), even if academic studies find Google Search can help forecasting, the ultimate 'replication' would be that there are businesses and government that actually use this knowledge in real life, and because of this, have a competitive advantage. That would be the 'real' proof that the predictive power of Google search intensity, is not just an academic gimmick but does provide an economic meaningful advantage to forecasters. [Mui \(2014\)](#) writes that 'The Bank of Israel and the Bank of England incorporate Google analytics into some of their forecasts.', suggesting adding Google Search intensity might even pass this advanced replication benchmark. At least, in some cases, that is.

---

<sup>6</sup> Additionally, one can also vary several dimensions at the same time.

## References

Choi, H. and H. Varian (2012). “Predicting the Present with Google Trends”, *The Economic Record*, Vol. 88, Special Issue, pp. 2–9.

Ettredge, M., Gerdes, J. and Karuga, G. (2005). “Using Web-based Search Data to Predict Macroeconomic Statistics”, *Communications of the ACM*, vol. 48 (11), pp. 87–92.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M. and Watts, D.J. (2010). “Predicting Consumer Behavior with Web Search”, *Proceedings of the National Academy of Sciences*, vol. 7 (41), pp. 17486–90

Mui, Q. (2014). “What search engines tell governments about the economic here and now”, *The Washington Post*, June 14, 2014.

Tuhkuri, J. (2016). “Forecasting Unemployment with Google Searches”, *ETLA Working Papers 35*, The Research Institute of the Finnish Economy.

Table 1: the contribution of search intensity series to the prediction of car sales under various scenarios

	Choi-Varian (2012) paper	Replication of Choi- Varian (2012) using their data and code	Replication of Choi- Varian (2012) using my data and code	Replication of Choi- Varian (2012) using my data and code and longer period (Jan 2004-July 2017)	Replication of Choi- Varian (2012) using my data and code for NZ	Replication of Choi- Varian (2012) using my data and code for NZ, and longer period (Jan 2004-July 2017)
	(I)	(II)	(III)	(IV)	(V)	(VI)
Intercept	-0.45798	-0.45798	-0.6818	0.4856	1.5787*	-0.5113
lag(y, 1)	0.61947***	0.61947***	0.5298***	0.4955***	0.2984***	0.4002***
lag(y, 12)	0.42865***	0.42865***	0.3565***	0.4027***	0.3973***	0.5893***
Trucks & SUVs	1.05721***	1.05721***	0.95***	0.336***	0.0243	0.2159*
Auto Insurance	-0.52966***	-0.52966***	-0.5098***	-0.1899***	0.2428	-0.0647
Adjusted R2 with search series	0.808	0.808	0.783	0.864	0.528	0.772
Adjusted R2 without search series	0.7111	0.7111	0.714	0.847	0.482	0.767

\*\*\* means significant at 1% significance level, \* means significant at 10% significance level.



Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

<http://www.economics-ejournal.org/economics/discussionpapers/2017-76>

The Editor