



January 2013

Does Google Search Data Aid In Predicting Unemployment?

Brennan Dyk

Follow this and additional works at: <https://commons.und.edu/theses>

Recommended Citation

Dyk, Brennan, "Does Google Search Data Aid In Predicting Unemployment?" (2013). *Theses and Dissertations*. 1415.
<https://commons.und.edu/theses/1415>

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.

DOES GOOGLE SEARCH DATA AID IN PREDICTING UNEMPLOYMENT?

by

Brennan Nicklaus Dyk
Bachelor of Arts, University of North Dakota, 2013

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

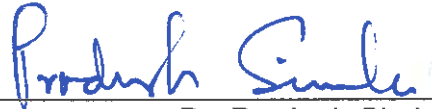
Master of Science

Grand Forks, North Dakota
August
2013

This thesis, submitted by Brennan N. Dyk in partial fulfillment of the requirements for the Degree of Master of Science in Applied Economics from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.



Dr. Cullen F. Goenner, Chairperson



Dr. Pradosh Simlai

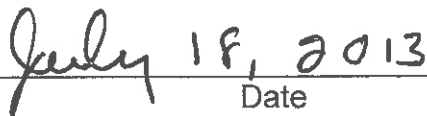


Dr. Daniel Biederman

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the Graduate School at the University of North Dakota and is hereby approved.



Dr. Wayne Swisher



Date

PERMISSION

Title Does Google Search Data Aid in Predicting Unemployment?
Department Applied Economics
Degree Master of Science in Applied Economics

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Brennan Dyk
August 1, 2013

TABLE OF CONTENTS

LIST OF FIGURES.....	v
LIST OF TABLES.....	vi
ACKNOWLEDGMENTS.....	vii
ABSTRACT.....	viii
CHAPTER	
I. INTRODUCTION.....	1
II. BACKGROUND.....	4
III. FORECASTING BUSINESS CYCLES & UNEMPLOYMENT.....	7
IV. DATA AND METHODOLOGY.....	11
V. RESULTS AND INTERPRETATION.....	16
VI. CONCLUSION.....	20
REFERENCES.....	23

LIST OF FIGURES

Figure	Page
1. United States Unemployment from January 2004 until May 2013.....	5

LIST OF TABLES

Table	Page
1. Mean-square-error from out-of-sample forecasting prior to July 2009.....	16
2. Mean-square-error from out-of-sample forecasting after April 2009 to April 2010.....	17
3. Mean-square-error for expansionary and full periods from December 2008 to August 2009.....	18
4. Mean-square-error for expansionary and full periods from June 2012 to May 2013.....	19
5. Mean-square-error for full predictive sample from July 2007 to May 2013.....	19

ACKNOWLEDGMENTS

I would like to thank all of my professors and particularly my thesis chair, Cullen Goenner. I would also like to thank my classmates Whitney Morsching, Doug Hart, Nathan Orwick, Yi (Julia) Liu for all of their support and encouragement. I would also like to thank my dad for being very supportive throughout my college career.

ABSTRACT

The unemployment rate is typically forecasted in the literature using initial claims for unemployment, a lagged value of unemployment, and an autoregressive component. This paper looks to improve upon existing models by adding Google search data to traditional models using initial claims or replacing initial claims with Google searches. One hypothesis is that Google searches may improve forecast accuracy due to employees knowing or getting a sense when they may become unemployed and be searching for jobs prior to filing for unemployment. Study of this is important to improve the accuracy of models and to provide more accurate out-of-sample forecasts for policymakers. If it showed a significant increase, it could perhaps diminish the frequency of surveys conducted by the Bureau of Labor Statistics. Several ARIMA models are used to determine whether or not the addition of specific Google searches can be more useful in an out-of-sample predictive model of unemployment. The results show that while Google was a good way for predicting the model in the past, it does not beat traditional models that use initial claims as an independent variable in predicting changes in the direction of the unemployment rate. In addition, models including regime switching do not improve the forecast accuracy compared with standard models.

CHAPTER I

INTRODUCTION

One of the largest determinants to policymakers for assessing the health of the economy at both a national and state level is the unemployment rate. Since the unemployment rate is reported in the first week after the end of the month, there is a lag in the direction of the unemployment rate. Policymakers and businesses would be better able to get a sense of where the country is going by forecasting the value ahead of time, or even the current state using nowcasting. Due to the lag, the unemployment rate is usually reported every single month about a week after the end of the month based on a monthly survey conducted by the Bureau of Labor Statistics (BLS). Traditionally, models have used initial claims for unemployment as a way to help with forecasting and nowcasting in both monthly and quarterly predictions.

When the economy declines or a worker knows that they are about to become unemployed, they may start searching for either unemployment benefits or potential jobs on a search engine such as Google. The definition of unemployed workers is workers who are not employed but are actively searching for work. The people that are searching for jobs on Google are by definition searching for work, and thus will not fall in the category of not in the labor force. Those searching for unemployment benefits will be forced to keep looking for

jobs as that is a requirement of those that are going to be unemployed. Although these searches on Google may seem trivial in the scope of all searches, the next paragraph will show how they have been good indicators in the past.

As we continue into the digital age, more and more data are becoming available. There are data from surveys, social media, and search engines that is available to researchers. One particular trend that researchers are using is data from Google searches to attempt to predict certain variables of interest. One recent application was the prediction of movie popularity in the box office (Google Whitepaper, 2013). Another practical application is in predicting influenza outbreak (Ginsberg et al., 2009). Google has also been a good indicator in predicting monthly sales figures for certain industries (Choi & Varian, 2012). With the spike in popularity of using Google as a forecasting variable, this paper looks to see if Google can be useful as a way to predict unemployment.

Since the unemployment rate is based on a monthly survey collected by the BLS, the methodology should be similar to the way that they collect their data. The time frame is similar to that of the BLS and this paper attempts to determine changes in those that are out-of-work using changes in Google searches on relevant search topics.

Although the initial claims number is a commonly used forecast variable, it begs the question: is there another variable to use with it or a better variable to

replace it? To answer the question, this paper will be looking at modeling over longer time periods than D'Amuri and Marcucci's (2010) paper as well as employing regime switching models in periods of expansion and contraction. The potential significance of using Google data is that it could be better and may be able to improve accuracy in forecasting and nowcasting. If the accuracy could be increased enough, it might be possible to reduce the frequency of the monthly sampling of the unemployment rate to save money. In addition, it would help provide a better idea where the economy is going on a timelier basis. This paper will examine different types of Google searches to see if they can be a useful tool to aid in the prediction of the United States unemployment rate.

CHAPTER II

BACKGROUND

The United States was recently devastated by the Great Recession in 2008. The Great Recession was characterized by rapidly increasing unemployment as shown in Figure 1 below. In 2009, the United States unemployment rate reached 10%, a level not seen since the 1980's and the Great Depression Era. In November 2007, the US real GDP forecast for 2008 was an increase of 2.7% in real GDP, which was revised down from 3.1% (Kaiser & Bohan, 2007). Instead of real GDP increasing by 2.7%, actual GDP for the year decreased by 4.2% (Federal Reserve Bank of St. Louis, 2013). This particular recession lasted from December 2007 until June 2009.

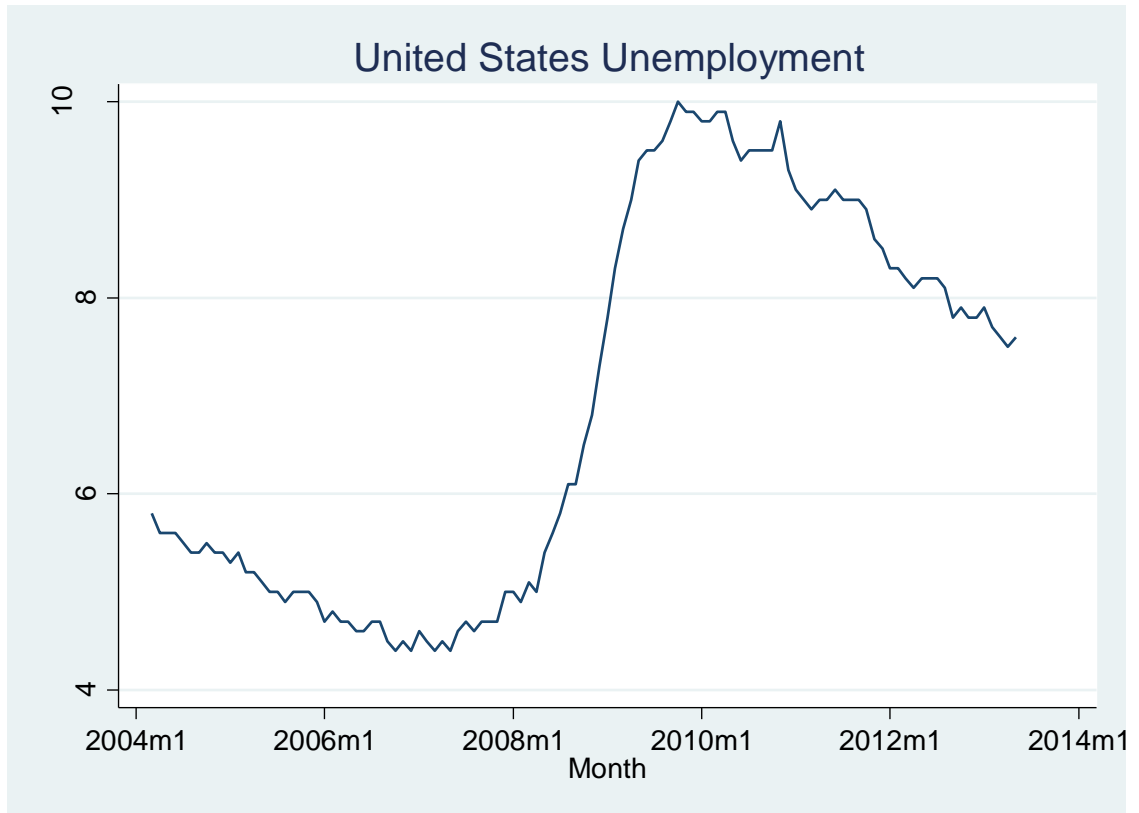


Figure 1. United States Unemployment from January 2004 until May 2013.

The reason that forecasting the unemployment rate is important is that it can help show the current and future state of the economy. This is important in determining trends and predicting turning points. One thing that is evident from Figure 1 is the peak of unemployment in 2009 was followed by a steady decline. Being able to predict those turning points is something that would be of significant interest in a model. Since the unemployment rate seems to be leveling off, the next direction that it takes becomes more and more important. In a time of uncertainty, it is very important from a policymaker's perspective to accurately

model and forecast unemployment to know where the future is going so that they can respond correctly using contractionary or expansionary policies.

The idea of using Google search data to forecast unemployment is not entirely unique, but this paper follows the ARIMA (1,1,0) models similar to previous papers (Montgomery, Zarnowitz, Tsay, & Tiao, 1998) (Marcucci & D'Amuri, 2010) (Rothman, 1998). The ideas and methodologies are applied to a more recent dataset than previous papers. Since the data for Google only goes back to January 2004, this provides some potential problems for limited samples and sample bias in previous work and this paper as well. This paper will look at using a method similar to D'Amuri and Marcucci's model (2010) in their time period, a full sample, and a piecewise regime switching model. The results show that models without Google do the best job with this available data.

CHAPTER III

FORECASTING BUSINESS CYCLES & UNEMPLOYMENT

While there has been some work published on predicting unemployment rates, much of it has been on the types of models to use and the different potential forecast variables. In looking for meaningful ways to predict unemployment, this paper considers unemployment forecasting as well as related forecasting of the business cycle.

Montgomery, Zarnowitz, Tsay and Tiao (1998); Emery and Koenig (1992); and Fritsche and Kouzine (2005) all suggest using Markov regime switching models to pool together times of expansionary and contractionary periods. Montgomery, Zarnowitz, Tsay and Tiao use a method that defines a period that had a negative unemployment of .3 percentage points as a down-period until it reached an up-period by moving up .1 percentage points or more. Predicting these periods is also important to be able to know which part of the switching model to use when predicting unemployment.

Because of using different models, one thing that is of particular interest is being able to predict when the economy is in expansion and when it is in a contraction. The point at which the economy changes from expansionary to contractionary is known as a turning point. The literature suggests that better models might be created when using only an expansionary or only a contractionary period. However, when applying the models to a turning point,

they often have higher variance and can struggle at those turning points. (Emery & Koenig, 1992)

Another approach for estimating a turning point is suggested by Chin, Gweke, and Miller (2000). They used the 12 month lag as a reference point as well as the two months around the particular month. The reason they used their own definition, not the National Bureau of Economic Research's (NBER) definition, is to be timely and consistent in their modeling. The NBER does not acknowledge turning points until well after the fact. For example, the last turning point was June 2009, but was not acknowledged until September 2010. Chin, Gweke, and Miller also use their definition of a recession as the dependent variable in their recession probability forecasting using a probabilistic model. In predicting turning points, commonly used variables include the manufacturing capacity utilization rate, the spread of Moody's AAA bonds to the 90-day treasury rate, initial claims for unemployment, and manufacturers' orders among many things (Chin, Gweke & Miller, 2000) (Negro, 2000).

Although predicting turning points is very important due to the potential difference in contractionary and expansionary models, it only helps us get an idea of the complexities of forecasting unemployment. Upon reviewing the literature specifically about forecasting the US unemployment rate, one finds that the US unemployment rate is non-linear and cannot be modeled with standard ordinary least squares. This is supported by Montgomery, Zarnowitz, Tsay and Tiao (1998), Golan and Perloff (2004), and Rothman (1998). Instead of using the standard OLS model, the aforementioned approach with either an autoregressive

model or a non-parametric method. The autoregressive modeling continues in use by D'Amuri and Marcucci (2010).

Another issue that has been suggested by the literature with forecasting the unemployment rate includes a unit-root of the data shown by the Dickey-Fuller test. Because of its presence, different transformations are needed. Rothman (1998) used a log-linear model to transform the data in order to remove the unit root. The commonly accepted method of removing the unit root is to use first differencing which is employed by almost all of the literature including Montgomery, Zarnowitz, Tsay and Tiao (1998); Chin, Geweke, and Miller (2000); and D'Amuri and Marcucci (2010).

When looking to forecast the unemployment rate using Google, there are many ways to predict. Zimmermann and Askitas (2009) model unemployment in Germany using search terms translated to be "unemployment office" or "unemployment agency" and "most popular job search engines in Germany". D'Amuri and Marcucci (2010) attempt to model the unemployment rate of the United States by using the search term "jobs". Lastly, Choi and Varian (2012) use two indices created by Google to model initial claims for unemployment. The first is a jobs index, which is created from searches related to job searching. The other is an unemployment index, which is created from searches related to unemployment and unemployment benefits.

How might Google searches add more information to simply initial claims? One reason that Google searches could be so enticing is the fact that they are published on a daily basis. This fact alone makes it much more timely than

standard models utilizing initial claims. Google also might provide information about those who know there are going to be without work in the short time to come, but may begin searching for jobs right away.

Zimmermann and Askitas (2009) used 65 data points in the time period from January 2004 to May 2009 and cautioned about the potential sampling bias. One of the disadvantages of their paper is they do not provide out-of-sample predictions to compare. Looking at the data from D'Amuri and Marcucci, they used a very short time period, 38 months, when employing lags from the first month in 2004 to the second month in 2007 and use March 2007 to June of 2009 as a testing period for the model. They find that their best model in terms of MSE comes in the model using both Google searches and initial claims for unemployment. The analysis in this paper looks to expand on this with more recent data.

CHAPTER IV

DATA AND METHODOLOGY

Before gathering the data, it's important to know how the unemployment data that is released every month is collected. The Bureau of Labor Statistics conducts a survey of 60,000 households every month during the week that includes the 12th of the month (Bureau of Labor Statistics, 2009). Since we want to model the dataset that is produced from the survey, this time period could be very important because it tells us when to look for Google search data and initial claim data. To attempt to replicate it, we will specifically look for Google and initial claims data from week containing the 12th of the month and the week prior.

The initial claim dataset is freely available from the United States Department of Labor in weekly format (United States Department of Labor). There seems to be about a two week lag from their website database, so the most recent data are available from their press releases.

When gathering information about Google searches, there is a lot more flexibility in this area. With Google, one is able to extract information about any search back to 2005 using Google's website called Google Trends¹. Google generally quotes the index with a maximum reference point. When searches for the term "jobs" were at their highest, the value will be 100. Every other

¹ The data is available from Google's website at <http://trends.google.com>.

observation is normalized relative from that reference point to be lower from 0 to 99. The search term that I decided to use from this method was the keyword “jobs” which was recommended by D’Amuri. Due to this being a relative index, it will differ my data from D’Amuri and Marcucci’s data due to rounding and there being less pronounced increases and decreases. Using a Granger test for causality, this keyword can be used to predict United States unemployment.

The next two possibilities that I tried were taken from an index that Google has built themselves. The dataset is taken from a webpage on Google Finance called “Google Domestic Trends²”. On it, they have built indices to be used for researchers based on searches that may be relevant to the index. Each index is standardized to 1 at its inception and can go up or down relative to that. The two that I chose that may be of interest would be their Jobs index and their Unemployment index.

When we look at the Unemployment index, the top terms that are used in calculating it are "unemployment", "food stamps", "social security" and "unemployment benefits". By running the Granger test for causality on the Unemployment index first, it shows us that the Unemployment index is not a suitable independent variable and that the US unemployment rate better explains the Unemployment index. This makes sense as when people hear about higher unemployment, they may do a Google search for unemployment to either read more about it or to find out the unemployment rate. The unemployment index

² The data is available from the Google Finance website at https://www.google.com/finance/domestic_trends.

may be the dependent variable while the release of the unemployment rate may be the independent variable.

When looking at the Jobs index, the top terms that are used in calculating it are “jobs”, “job”, “salary”, “resume”, and “careers” to name a few. Running the Granger causality test on the Jobs index, we find that this is a much more suitable independent variable to use than the unemployment index. This also makes sense because we don’t expect people to search for jobs after finding out about low or high unemployment. We would expect the searches for jobs to be independent of the release of unemployment numbers. We expect that the amount of people searching for jobs will be indicative of the unemployment rate to be released.

When using the unemployment data, I initially used not seasonally-adjusted data due to the similar nature of the data from Google. When using monthly dummy variables, the regression will give you higher levels of significance than it actually provides. Because of that, I realized that a seasonally-adjusted dataset provides better information in terms of increases or decreases and provides smoothing which should help build models with more meaningful results. The Google dataset was deseasonalized using a monthly multiplicative factor.

Another consideration to look into before modeling is whether or not the unemployment set is affected by a unit root. If there is a unit root, it means there are autocorrelated residuals. A consequence of that is that we will have bias with higher levels of R^2 and meaningless coefficients (Granger & Newbold, 1974) .

Using a Dickey-Fuller test, we find that there is indeed a unit root problem since we are unable to reject the null of a unit root at a p-value of .7526. This problem can be alleviated by first-differencing. Using the standard that was followed in the literature review, they do use first-differencing when using ARIMA (1,1,0) models. This paper will also be following the standard ARIMA (1,1,0) model.

The goal in any model is to be able to be as accurate as possible. However, in this context, a more important determinant than in-sample fitting will be to see how well it can do out-of-sample. If we add a ton of variables to the model, it will probably fit the in-sample data better, but will not necessarily make better out-of-sample predictions. The goal here is going to be to be able to add variables that will improve our out-of-sample forecasting. The models in this paper will focus on using rolling samples so that it can predict in the same way that a forecaster without any knowledge of the future might predict using past variables.

In the models, the traditional base case of using initial claims for unemployment is compared against models including a Google variable. The models that are attempted vary. One type of model includes initial claims in addition to jobs searches. The other model that is attempted contains only a Google independent variable. D'Amuri and Marcucci use lagged values of monthly and weekly averaged data to predict ahead into the future. It is important to look to see if the inclusion of Google searches will provide better out-of-sample forecasting than those just using initial claims for unemployment. The model

specification that is used here is ARIMA (1,1,0) with independent variables of the week of the 12th of the month, the week prior, and two lags of each variable.

Since D'Amuri and Marcucci's sample time period was right before the first uptick in the unemployment rate, perhaps there is some bias that their models follow trends and cannot predict changes in the business cycle. In order to test for that, this paper will look at model accuracy during a turning point as well as comparing results to regime switching models.

CHAPTER V
RESULTS AND INTERPRETATION

In order to show that the model specification works, we need to first confirm that the model specification can replicate D’Amuri and Marcucci’s quantitative results, which say that models including Google data can lower the mean-square-error compared to models without it. The model is run on the same time period using similar data. Adding “jobs” searches to the traditional initial claims model, the new model reduces the out-of-sample MSE, in the time period March 2007 to July 2009, from .0473 to .0415, a reduction of about 12.2%. The other model reduces it from .0473 to .0441, a reduction of about 6.8%. This verifies the quantitative result that during their time period, Google search results can reduce the error.

Table 1. Mean-square-error from out-of-sample forecasting prior to July 2009.

	Initial Claims (IC)	“Jobs”	“Jobs” and IC	Job Index	Job Index and IC
MSE, lags 0-2	0.04733	0.07625	0.04153	0.09146	0.05040
MSE, lags 1-2	0.04733	0.07913	0.04411	0.08735	0.04453

There are two major problems with the first time period. The first is that the time period at which they ended was almost exactly at the turning point. That means that the part they built the model on was almost strictly going down, and the part where they predicted was almost strictly increasing. Because of the

nature of the data, OLS models tend to outperform the autoregressive models in this time period. If it is a good model, a turning point should not be an issue. To test its significance, it is more important to look at different periods of time. It is of particular interest to see how well the model does during a turning point. This paper identifies October 2009 as a turning point, and will look at a 6 month period on both sides from April 2009 to April 2010 to compare the out-of-sample MSE to see which model performs best.

In using the rolling models around the turning point, Table 2 shows that the initial claims model does the best, although not by a lot. This shows that Google does not necessarily do a better job forecasting compared to traditional models. In every case, the model of the initial claims slightly beats models including Google.

Table 2. Mean-square-error from out-of-sample forecasting after April 2009 to April 2010.

	Initial Claims (IC)	“Jobs”	“Jobs” and IC	Job Index	Job Index and IC
MSE, lags 0-2	0.03097	0.03635	0.04844	0.03223	0.03662
MSE, lags 1-2	0.01254	0.02910	0.01524	0.02707	0.01312

Although we have evidence against the idea that Google can aid in predicting turning points in this type of model, the last important result to look at is the full-sample model. The last comparison to test is whether or not the switching model will outperform the standard full-sample model. The idea here is to model the points of increase together and the points of decrease together. The

only difficulty here is the limited sample size that only allows for one area of economic decrease and two areas of economic increase.

Although it appears from the simple model here that Google doesn't really improve predicting in the longer term, it is important to remember the literature review. It has suggested that there are potentially better models to be created when they "switch" between contractionary and expansionary periods. That implies that building a model should consist only of periods where it is going up or down. The difficulty in forecasting is to determine that particular point where the model changes from going up to down.

In this sample, the contractionary period is relatively small to the expansionary period. What I have done is looked near the end of the contractionary period in order to maximize the amount of data needed to build a model. The results in Table 3 show that neither model outperforms the other in comparison. A similar result here is that initial claims again beat other models.

Table 3. Mean-square-error for expansionary and full periods from December 2008 to August 2009.

	Initial Claims (IC)	"Jobs"	"Jobs" and IC	Job Index	Job Index and IC
Contractionary MSE (lags 0-2)	0.010345	0.032757	0.028849	0.056724	0.105489
Full MSE (lags 0-2)	0.010577	0.061804	0.015281	0.056999	0.013404
Contractionary MSE (lags 1-2)	0.014026	0.027842	0.022979	0.063848	0.07458
Full MSE (lags 1-2)	0.017263	0.049417	0.017881	0.067192	0.022099

When we apply the switching model to the expansionary time period using the largest sample, we find the results shown in Table 4 below. We find that the expansionary model on average below does worse in terms of model performance.

Another interesting note from Table 4 is that the initial claims model is highly ranked among the models in terms of MSE. This backs up the previous results in Table 2 and Table 3 that the initial claims dataset is still one of the best independent variables to use in predicting the unemployment rate.

Table 4. Mean-square-error for expansionary and full periods from June 2012 to May 2013.

	Initial Claims (IC)	“Jobs”	“Jobs” and IC	Job Index	Job Index and IC
Expansionary MSE (lags, 0-2)	0.02814	0.029446	0.036347	0.017543	0.024288
Full MSE (lags 0-2)	0.025419	0.004996	0.030976	0.016958	0.032772
Expansionary MSE (lags 1-2)	0.013935	0.013814	0.012689	0.015844	0.015016
Full MSE (lags 1-2)	0.01699	0.024289	0.017869	0.013105	0.016131

Lastly, to verify that selection bias is not an issue when saying that initial claims data still does the best job, I ran a rolling regression from July 2007 to May 2013. The results below in Table 5 show that on average, initial claims does best job by beating the next best model by 6.9% and 4.3% respectively.

Table 5. Mean-square-error for full predictive sample from July 2007 to May 2013.

	Initial Claims (IC)	“Jobs”	“Jobs” and IC	Job Index	Job Index and IC
MSE (lags 0-2)	0.025794	0.04016	0.0300	0.037458	0.027583
MSE (lags 1-2)	0.026843	0.03858	0.0291	0.036427	0.028043

CHAPTER VI

CONCLUSION

Although it seems that people searching on Google for jobs might be a good way to help predict unemployment, this paper has shown that it isn't better than the traditional initial claims model. When comparing to D'Amuri and Marcucci's paper, I found that in their time period, they were able to show that jobs provided a better model due to the Google data being more correlated with unemployment. In every other time period, we find that initial claims consistently performed the best or close to the best. Using the full sample test to eliminate selection bias, we find that initial claims on average does the best job in out-of-sample forecasting. In addition, the best models were those that included the full sample, rather than just the expansionary periods and contractionary periods.

When using Google data, there are many issues that could have potentially been problems. One problem may have been our sample. Other researchers use 40-50 years of data when generating models based on initial claims. Another potential issue is sample bias: is there a certain group of people more likely to use Google to search for jobs or use these particular job terms. The third potential problem is: do people still need Google to find the websites where the jobs are posted? If they have already found a website such as indeed.com or monster.com, do they need to keep using Google every time they

become unemployed? A fourth potential concern is changes in search engines. Are there more or fewer people using Google as their search engine when they are searching for these keywords? A fifth concern is that the people searching on Google are not unemployed, but are rather searching for a new job from their existing job. The aforementioned issues do not affect initial claims. The area where initial claims may have the largest error in predicting unemployment would be if hiring picks up, then those initial claims will not translate into higher unemployment.

Although I can't answer all the questions, we can check to see if Google usage has remained consistent from 2008 to the present (StatCounter GlobalStats). What we find from the monthly data is that Google usage in the United States has ranged from a low of 76% to a high of 82%, which is relatively in the same range. This implies that a problem is that people are using Google in particular less and less in searching for jobs.

Although there are some potential problems, there are some ways that the model could be improved in future study of the US unemployment rate. One might be to try to using Google to supplement a probit model to forecast recessions. The reason that this is not feasible now is that there is only one recession in the sample. Although the switching models did not turn out well, a future application might be to include a probit model to supplement the regime switching models to know when there might be a turning point in the data. Another idea, if it were available, would be to use the number of active users or page views on a website like monster.com or indeed.com to be an independent

variable of unemployment. This may be a more accurate measure of future unemployment without as much ambiguity of a search engine.

While the dataset from Google does not allow us to build a great model with the standard ARIMA (1,1,0) for unemployment, there may be other areas that the Google dataset does well such as predicting sales or public awareness for a company. Another application could be macroeconomic data other than the US unemployment rate. Although this research didn't show the unemployment rate to be a good variable, it does not rule out future research with this tool.

REFERENCES

- (n.d.). Retrieved June 20, 2013, from StatCounter GlobalStats:
http://gs.statcounter.com/#search_engine-US-monthly-200807-201305
- Bureau of Labor Statistics. (2009, October 16). *How the Government Measures Unemployment*. Retrieved June 29, 2013, from Bureau of Labor Statistics:
http://www.bls.gov/cps/cps_htgm.htm
- Chin, Geweke, & Miller. (2000). *Predicting Turning Points*. Minneapolis: Federal Reserve Bank of Minneapolis.
- Choi, & Varian. (2012). Predicting the Present with Google Trends. *The Economic Record* , 88, 2-9.
- Emery, & Koenig. (1992). *Is a Two-State Characterization of the Business Cycle Appropriate?* Dallas, TX: Federal Reserve Bank of Dallas.
- Federal Reserve Bank of St. Louis. (2013). *Real Gross Domestic Product*. Retrieved July 1, 2013, from Federal Reserve Economic Data:
<http://research.stlouisfed.org/fred2/series/GDPC1>
- Fritsche, & Kouzine. (2005). Prediction of Business Cycle Turning Points in Germany. *Journal of Economics and Statistics*. , 225 (1), 22-43.
- Ginsberg. (2009). Detecting influenza epidemics using search engine query data. *Nature* (457).
- Google Whitepaper. (2013, June). *Google Think Insights*. Retrieved June 2013, from Google: <http://www.google.com/think/research-studies/quantifying-movie-magic.html>
- Granger, & Newbold. (1974). Spurious Regressions in Econometrics. *Journal of Economics* , 2 (2), 111-120.
- Kaiser, & Bohan. (2007, November 29). *White House cuts 2008 economic forecast*. Retrieved July 1, 2013, from Reuters:

<http://www.reuters.com/article/2007/11/29/us-bush-economy-idUSWBT00798620071129>

Marcucci, & D'Amuri. (2010, April). "Google it!" *Forecasting the US Unemployment Rate with a Google Job Search index*. Retrieved May 30, 2013, from Social Science Research Network:
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1594132

Montgomery, Zarnowitz, Tsay, & Tiao. (1998). Forecasting the U.S. Unemployment Rate. *Journal of the American Statistical Association* , 93 (442), 478-493.

Negro, M. d. (2000). *Turn, Turn, Turn: Predicting Turning Points in Economic Activity*. Atlanta: Federal Reserve Bank of Atlanta.

Perloff, & Golan. (2004). Superior Forecasts of the U.S. Unemployment Rate Using a Nonparametric Method. *The Review of Economics and Statistics* , 86 (1), 433-438.

Rothman, P. (1998). Forecasting Asymmetric Unemployment Rates. *The Review of Economics and Statistics* , 80 (1).

United States Department of Labor. (n.d.). *Unemployment Insurance Weekly Claims Data*. Retrieved June 1, 2013, from United States Department of Labor:
<http://www.ows.doleta.gov/unemploy/claims.asp>

Zimmermann, & Askitas. (2009). Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly* , 55 (2), 107-120.