Title
Three Essays on Big Data in International Finance

Author
Zang, Ziqi

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Three Essays on Big Data in International Finance

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Ziqi Zang

2019

ABSTRACT OF THE DISSERTATION

Three Essays on Big Data in International Finance

by

Ziqi Zang

Doctor of Philosophy in Economics

University of California, Los Angeles, 2019

Professor Aaron Tornell, Chair

This dissertation presents an introduction to big data that can potentially be used in nowcasting key macroeconomic variables for advanced economies. It also explores the forecastability of big data in short-term exchange rate forecasting. Finally, it draws on evidence from a sentiment analysis of Article IV Consultations over the period of 2012 to 2018 and examines the development of member countries' perceptions of IMF policy advice.

Chapter 1 uses big data from Google search data to form better nowcasts of macroeconomic variables. My empirical strategy contributes to the macroeconomic nowcasting literature on three fronts. First, I take a number of steps to identify the most comprehensive set of relevant search queries that capture people's search behavior in relation to each monetary policy variable, such as the unemployment rate and inflation. Second, I consider regularization and dimension reduction methods to handle the underlying high-dimensional regressor space with highly correlated covariates. Third, I evaluate both average point forecasts and conditional point forecasts against benchmark models with DMW test and CSPA test, respectively. According to the test statistics, I find that Google search data offer significant improvements in nowcasting macroeconomic variables both unconditionally and conditionally.

Chapter 2 examines the short-term forecastability of exchange rates using machine learning models in a rich data environment. I investigate the performance of different machine learning models, such as variable selection models, dynamic factor model, and decision regres-

sion trees in obtaining accurate forecasts of three currency pairs (U.S./U.K., Japan/U.S. and U.S./Australia). I consider three types of forecasts: point forecasts, unconditional weighted directional forecasts and conditional weighted directional forecasts. According to the DMW test, out-of-sample forecasts of every currency rejects the null hypothesis of equal forecasting errors with the random walk with at least one machine learning model. Furthermore, the conditional weighted directional forecasts allow us to know *when* exactly our models are more profitable than the random walk with zero profit. And it turns out that our weighted directional forecasts are significantly positive especially on the tails of the conditioning variable distribution.

Chapter 3 constructs multi-aspect policy sentiment measurements to interpret authorities' tones in response to specific policy advice in IMF Article IV Consultations. Specifically, we use a topic-based sentiment analysis approach that entails the application of a latent Dirichlet allocation (LDA) model as well as sentiment prediction machine learning models. Therefore, we are able to provide the stylized facts that provide useful input for assessing the impact of Fund advice on macroeconomic development of member countries.

The dissertation of Ziqi Zang is approved.

Zhipeng Liao

Pierre-Olivier Weill

Nico Voigtläender

Aaron Tornell, Committee Chair

University of California, Los Angeles

2019

Dedicated to my beloved parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Throughout the writing of this dissertation I have received a great deal of support and assistance.

First and foremost, I would like to to thank my advisor Professor Aaron Tornell and Professor Zhipeng Liao, whose expertise is invaluable in the formulating of my research topic and methodologies. I would also like to thank Professor Pierre-Olivier Weill and Professor Nico Voigtläender for their help.

I would like to acknowledge my colleagues at IMF for offering me a great internship opportunity. I would particularly like to thank my supervisor Dr. Vivek Arora for his great help and recommendation. I thank Marie-Therese Camilleri, Iryna Ivaschenko, and Xin Li for their expertise and assistance throughout writing Chapter 3.

I also thank my former manager Paul Rohwer, for introducing me to the data scicence world.

Last, I would like to thank my parents, whose unconditional love and support are with me in whatever I pursue.

VITA

2013	B.Sc, Economics (Minor in Mathematics) University of Minnesota, Twin Cities Minneapolis, Minnesota
2016	M.A., Economics University of California, Los Angeles Los Angeles, California
2016	C.Phil., Economics University of California, Los Angeles Los Angeles, California

CHAPTER 1

Macroeconomic Nowcasting with Big Data

1.1 Introduction

Real-time monitoring of macroeconomic conditions has become important for policy making at central banks. Moreover, releases of macroeconomic data that turn out to be surprises will move markets, and investors adjust their expectations about the current state of the economy while relying on these releases. However, most macroeconomic data are released with a time lag and subsequently revised. A growing body of studies on macroeconomic nowcasting uses pervasive available big data to provide timely estimates or proxies of macroeconomic variables weeks or months before official releases. And Google search data has been increasingly incorporated in nowcasting or short-term forecasting of macroeconomic variables (Choi and Varian, 2009; Askitas and Zimmerman, 2009; Ginsberg et al., 2009; Baker and Fradk, 2017). Available since January 2004, Google Trends search queries can help forecast future data releases if people conduct an online search for information prior to making decisions. For example, when someone becomes unemployed or is laid off, it is expected that she or he will look for information such as "unemployment office," "unemployment benefits," "part time jobs," "resume," and other words related to finding a job. In this paper, I show that using Google search data can bolster our understanding of real-time macroeconomic developments.

How can we use big data from the search engine to form better nowcasts of macroeconomic variables? My empirical strategy makes contributions to the macroeconomic nowcasting literature on three fronts. First, I conduct a number of steps to identify the most comprehensive set of relevant seed queries that well capture people's search behavior in relation to each monetary policy variable, such as unemployment rate and inflation. Moreover, the seed queries are tailored for different countries because of the dramatic differences in cultures and social systems. For example, people might look for "redundancy pay" or "job seeker allowance", or register with a "JobCenter" when they are made redundant in the U.K., whereas people in the U.S. care about "unemployment claims" and "COBRA." After collecting the seed queries, I expand the seed query set empirically further, using semantic vectors. Then I feed these queries into Google Correlate to generate a 100 of the most correlated search queries with each of the seed queries. Unlike recent studies using Google search queries such as Bok et al. (2018), Bulut (2018), and Chojnowski and Dybka (2017), I take more transparent steps to select the set of predictors to be included in the following nowcasting task.

Second, I construct "nowcasts" for macroeconomic fundamental variables by selecting empirically relevant predictors from a set of high-dimensional Google search queries. I consider regularization and dimension reduction methods to handle the underlying high-dimensional regressor space with the covariates that are highly correlated. Specifically, I use the adaptive elastic-net model (Zou and Zhang, 2009) and form one-step-ahead out-of-sample forecasts relying on cross-validated estimators.

Third, I evaluate the forecasting ability of the model with Google Trends data against the benchmarks and consider two types of forecasting methods. Point forecast performance is based on the mean squared error (MSE). My results show that most of the point forecasts have significantly smaller MSE than those implied by the driftless random walk according to the Diebold Mariano and West (DMW) test. However, only two out of eight macroeconomic forecasts (unemployment rate of Australia and U.K. inflation) outperform an autoregressive model. However, the unconditional predictive test fails to answer *when* the model outperforms the benchmarks. Therefore, I consider the conditional superior ability test proposed by Li, Liao, and Quaedvlieg (2019) to assess the relative predictive performance of my model against the benchmarks conditional on a cyclical indicator variable (e.g., average inflation over the past months). And all of the point forecasts outperform both of the benchmarks, at least during some periods between 2009M4 and 2019M1, and especially during low inflation episodes.

Fourth, if the Fed and other central banks follow the Taylor rule to set their short-term nominal interest rates and my macroeconomic nowcasts can proxy the current states of the labor markets and prices, these search queries may demonstrate the forecastability of interest rate differentials. I show that these "collective wisdoms" have more than a 60% average success ratio of three pairs of one-month-ahead interest rate differential forecasts. Furthermore, I construct a simple exchange rate forecasting strategy based on the "surprises" from my directional interest rate differential forecasts. I forecast appreciation (depreciation) in U.S. dollars between t and $t+h$ ($h = 1, 3, 6, 9$, and $12m$) if there is a forecasting surprise of increase (decrease) in the interest rate differential. In the other cases, I prefer random walk which predicts zero exchange rate change.

The remainder of this chapter is organized as follows. In Section 1.2 and Section 1.3, I provide a short literature review of macroeconomic nowcasting and Google search data. Section 1.4 describes the procedure I use o construct the potential predictors for macroeconomic variables from Google search data. Also, I describe a recursive-window nowcasting model with high-dimensional predictors. Section 1.5 presents and evaluates two types of forecasting measurement—unconditional and conditional point forecasts. In Section 1.6, I demonstrate an exchange rate forecasting strategy based on the "surprises" from interest rate differential forecasts. Section 1.7 concludes.

1.2 Related Literature

Macroeconomic data such as the unemployment rate and inflation, are typically published with a time lag. An emerging literature on nowcasting (Giannone, Reichlin and Small, 2008; Aruoba, Diebold, and Scotti 2009; Choi and Varian, 2009, 2011; Scott and Varian, 2013; Banbura, Giannone, and Modugno, 2013; Scotti, 2013; Koop and Onorante, 2013; Carriero, Clark and Marcellino, 2015; Tatjana, Guenette and Vasishtha, 2017) focuses on using big data for macroeconomic nowcasting. This research thus provides timely estimates of macroeconomic variables such as GDP, prices, inflation, and unemployment, and proposes big-data based leading indicators of economic activities.

This chapter contributes to the nowcasting literature by examining whether real-time internet search data improve nowcasting power beyond a conventional set of regressors. Several sources of data on real-time economic activity are now available, such as professional forecasts, surveys, newspapers, micro blogs, Internet searches, and many others. Choi and Varian (2009, 2011) claim that Google search data may help in "predicting the present." There is also am emerging body of research indicating that Google search data are potentially useful in nowcasting economic variables such as in Bok et al. (2017), and Koop and Onorante (2013). I provide a list of recent studies in Appendix 1.6. The choice of keywords is a crucial ingredient when using search data for prediction. However, the out-of-sample nowcastability in previous mentioned studies highly depends on a set of deliberately selected search queries. And they are silent about constructing their predictors or arbitrarily select queries according to in-sample estimates. In contrast, I provide a comprehensive procedure to collect the most relevant search queries that answer question, "what are people searching online when they face unemployment or worry about prices?"

Macroeconomic time series are typically short, yet there is an enormous number of relevant Google search queries. Traditional econometric models are not suitable when the set of predictors is much larger than the number of observations. Also, I expect a high degree of sparsity in the regressor space, in the sense that the coefficients for a vast majority of predictors will be zero. To deal with this high-dimensionality, many studies on macroeconomic forecasting use sparse modeling such as LASSO with l_1 penalty (Bai and Ng, 2008; De Mol et al., 2008; Marsilli, 2014; Nicholson et al., 2015; Uematsu and Tanaka, 2017). Although a l_1-penalty is expected to achieve good prediction accuracy, it fails to provide consistency in model selection. Moreover, when the regressor space is highly correlated, the model with only l_1 penalty will prevent group selection. To handle the collinearity problem properly in a rich data environment, Zou and Zhang (2009) introduce an adaptive elastic-net model that combines l_2 and l_1 penalization terms and considers the adaptive LASSO shrinkage to remedy the drawbacks of LASSO model.

Regarding the forecast evaluation method, the most commonly used method in empirical macroeconomics is the DMW test, which considers the unconditional equal predictive performance of two competing models. Under the null hypothesis, the forecasting models have the same expected forecasting errors. In addition to a conventional point forecast test, I assume further that the nowcasts may perform heterogeneously during expansion periods from recessions. Therefore, it is important to assess our forecasts conditional on certain economic states. Li, Liao, and Quaedvlieg (2019) propose a test of conditional superior predictive ability (CSPA) that tests inequalities for the conditional expectation functions of forecast errors. The null hypothesis of the CSPA test states that the conditional expected forecasting error of the benchmark is less than those of competing forecasts uniformly across all conditioning states. While under the alternative hypothsis, the alternative forecast method outperforms the benchmark in certain states. This paper applies the CSPA test to the point forecasts. And it turns out that my nowcasts perform better than the benckmarks especially in relatively low inflation periods.

1.3 Google Search Data

In this section, I demonstrate a series of steps to construct predictors from raw data by using Google tools. Section 1.3.1 shows how to obtain a wide pool of search terms from scratch to understand people's search behavior in relation to unemployment and inflation. In Section 1.3.2, I use Google Autocomplete to capture the nuances of how people search different topics. Google Correlate also helps to expand these to a high-dimensional set of predictors.

1.3.1 Search Term Selection Process

Google Search provides an excellent platform for observing people's information seeking activities. It offers immediate reflection of the demands and interests on the Internet. For instance, people who search for employment-related information, will type in "jobs" or "unemployment benefits." However, as we may expect, identifying all the possible ways the general public might search for a topic can be very difficult.

Luckily, I can make use of some Google tools to partially solve this problem, which allows us to identify the most comprehensive set of relevant search terms around a single topic. In this section, I develop a procedure that can help identify search terms and ensure solid representation of the possible ways people may search topics related to unemployment and inflation in different countries.

1.3.1.1 Seeding Search Queries

To start, I survey researchers from different countries of origin[1] to identify a pool of search terms that can generally characterize people's search behavior in different regions. During the sessions, I ask our researchers the following questions relevant to the labor market:

(i) What do you think people will look up online when they lose their jobs/are unemployed?

(ii) What are the unemployment protections/benefits/insurances in your country?

(iii) If applicable, how and where do people claim these?

For example, in the U.K., people who are made redundant have to register as unemployed with Jobcentre and are eligible for unemployment benefits such as Jobseekers Allowances (JSA) and Universal Credit. In Japan, people tend to go to Hello Work (ハローワーク), the Japanese government's employment service center, which manages unemployment insurance benefits (失業保険) for unemployed workers and also provides job-matching programs for the unemployed.

In terms of search terms relevant to inflation, Google uses its vast database of web shopping data to construct the "Google Price Index", which measures the U.S. inflation. Although I don't have the complete price history of items that measure the consumer price index, I hypothesize that search popularity of prices for relevant items conveys useful information to signal changes in inflation: The measure of price indices is based on a basket of consumer goods and services[2], and the set of search terms is based on people's interest in

[1] For help in developing some of this "inside baseball" knowledge, I thank Akina Ikudo (Japan), David Lindsay (Ireland), Carmen Rollins (Canada), and Lina Zhang (Australia).

[2] I exclude food and energy prices here, which are volatile for measuring inflation.

the prices of these items. To complete the list of search queries, I also consider and collect sentiment words such as "expensive" and "cheap" that people might use in response to rising or plummeting prices.

This results in the identification of around 30 possible search terms for each macroeconomic variable. I then narrow the list by testing each of these terms on Google Trends to estimate whether the popularity of each term makes sense (too volatile or too noisy). Also, I consider the casual and ingenious nature of people's use of search engine. For example, people may not have used the word "air transportation price" but rather "flight tickets."

1.3.1.2 Semantic Vectors

After selecting the seeding queries from researchers, I can further expand the keyword set empirically using semantic vectors. Unlike a word count method that relies on the distance between pairs of word vectors, Pennington et al. (2014) propose an unsupervised learning algorithm for obtaining word vector representations, called Global Vectors (GloVe). GloVe belongs to the famility of word embedding models. It leverages the advantages of both global matrix factorization methods and local context window methods to map words onto a latent vector representative space. In a comprehensive summary study by Gentzkow et al. (2017), this latent vector representative space, $\mathcal{V}$, contains the positions (word embeddings) for every vocabulary word in $I\!R^N$, where N is the dimension of the latent vector space. In a weighted least squares regression model, the word embeddings are determined in the loss function and can be interpreted as the likelihood of word cooccurences.

The GloVe word embedding model has been developed as an open-source project at Stanford. There are also pre-trained word vectors on various corpora, such as Wikipedia and Twitter data. I choose pre-trained word vectors that have been trained on 2 billion Twitter posts, with 1.2 million unique words. This choice of corpus allows us to understand the most frequent words people tweet, along with seeding queries such as "unemployed" and "unemployment benefits."

Table 1.1 lists the 50 most semantically related words people use in their tweets to the query “unemployed.” From the results, lots of the words, such as “jobless,” “poverty,” and “wages,” directly describe labor market condition. Similarly, Table 1.2 provides a set of semantic words to “inflation.” By leveraging the pre-trained word vectors, I can generate more queries people commonly use on social media that are analogous to the seeding queries. However, one weakness of using GloVe to expand the set of search queries is that it cannot output phrase-level semantic vectors.

1.3.2 Google Tools Utilization

Although I limit to search terms relevant to unemployment and inflation, these are broad terms that do not allow us to gain a more nuanced perspective on different motivations and interests of people’s search behavior, such as whether people are seeking information about “jobcentre login” or “jobcentre near me.” Considering this discrepancy, I use Google Autocomplete to generalize additional revelant terms that are often searched alongside the terms already entered. Figure 1.1 presents an example of using Google Autocomplete to expand relevant search terms with “jobcentre.”

Another way to expand the existing pool of search queries is by using Google Correlate. I feed each search query either from the list I brainstorm or Google Autocomplete, into Google Correlate. This will retrieve the set of individual search queries that are most highly correlated[3] with the input. Specifically, Google Correlate will provide 100 correlates for each input. Figure 1.2 is a screen shot of the tool, and shows the top 10 correlated queries to the search query of “jobcentre” and restricts the result in the U.K.

Through these steps, I identify over 900 relevant search queries for each macroeconomic variable after removing duplicates and spurious terms [4]. Then I remove seasonality and use first-differenced search data. See on online Appendix[5] for the full list of queries.

---

[3]Google Correlate surfaces queries in the database whose spatial or temporal pattern is most highly correlated (R^2) with a target pattern. Google Correlate uses an approximate nearest neighbor (ANN) algorithm over millions of candidate queries to produce results.

[4]I keep the ones that have plausible economic meaning. For example, some queries from Google Correlate are egregious anomalies that do not contain justifications.

[5]`https://github.com/zikiki/Dissertation-ONLINE_APPENDIX`

Table 1.1: Semantically Related Words to "Unemployed" (Top 50)

Ranking	Word	Semantic Similarity	Ranking	Word	Semantic Similarity
0	jobless	0.848683	25	increasing	0.626572
1	rates	0.756153	26	costs	0.625147
2	inflation	0.726879	27	average	0.625088
3	employment	0.723951	28	gdp	0.623232
4	incomes	0.717781	29	decrease	0.622389
5	wages	0.710747	30	adjusted	0.621555
6	joblessness	0.694542	31	benefits	0.615683
7	income	0.689244	32	employers	0.614083
8	rise	0.686522	33	drop	0.612629
9	jobs	0.685400	34	steady	0.611067
10	wage	0.679390	35	recession	0.610931
11	increase	0.677870	36	estimated	0.609868
12	percentage	0.673880	37	borrowing	0.604569
13	increases	0.672206	38	demand	0.601194
14	rising	0.666984	39	numbers	0.600426
15	decline	0.662284	40	housing	0.599370
16	lowest	0.661281	41	payments	0.598942
17	percent	0.658998	42	deficit	0.596879
18	growth	0.658234	43	workforce	0.593455
19	spending	0.642712	44	interest	0.590735
20	higher	0.639059	45	labor	0.590393
21	low	0.638412	46	risen	0.588437
22	poverty	0.636035	47	proportion	0.588003
23	economy	0.632944	48	premiums	0.587301
24	increased	0.630377	49	months	0.587176

Note: This table ranks the relatedness of the words to the seeding query "unemployed" according to their semantic similarity. In the pre-trained word vectors from Twitter, each word is represented by a word embedding vector w with a dimension 1×100. Semantic similarity is calculated by the product of the two word vectors, namely, $w_{unemployed}w_j^T$.

Table 1.2: Semantically Related Words to “Inflation Rate” (Top 50)

Ranking	Word	Semantic Similarity	Ranking	Word	Semantic Similarity
0	rates	0.900690	25	steady	0.672901
1	unemployment	0.782038	26	consumption	0.669353
2	rise	0.779551	27	inflationary	0.668337
3	growth	0.778871	28	deficit	0.667306
4	increases	0.756130	29	economy	0.665877
5	rising	0.741188	30	spending	0.662100
6	interest	0.740285	31	lending	0.661963
7	borrowing	0.735184	32	hike	0.660520
8	prices	0.731652	33	deflation	0.659799
9	higher	0.721807	34	percentage	0.659228
10	drop	0.715758	35	pressures	0.657168
11	decline	0.714473	36	hikes	0.653732
12	low	0.710563	37	decrease	0.653614
13	price	0.706645	38	consumer	0.649915
14	increase	0.703758	39	increasing	0.646997
15	gdp	0.699554	40	lower	0.646282
16	fed	0.689308	41	recession	0.642134
17	slowing	0.686638	42	percent	0.640819
18	yields	0.683581	43	slowdown	0.638271
19	expectations	0.683528	44	easing	0.632930
20	increased	0.682322	45	risen	0.632644
21	trend	0.677615	46	currency	0.628563
22	demand	0.676973	47	raise	0.628412
23	lowest	0.674087	48	costs	0.628325
24	levels	0.673124	49	benchmark	0.627972

Note: This table ranks the relatedness of the words to the seeding query “inflation” according to their semantic similarity. In the pre-trained word vectors from Twitter, each word is represented by a word embedding vector w with a dimension 1×100. Semantic similarity is calculated by the product of the two word vectors, namely, $(w_{inflation} + w_{rate})w_j^T$.

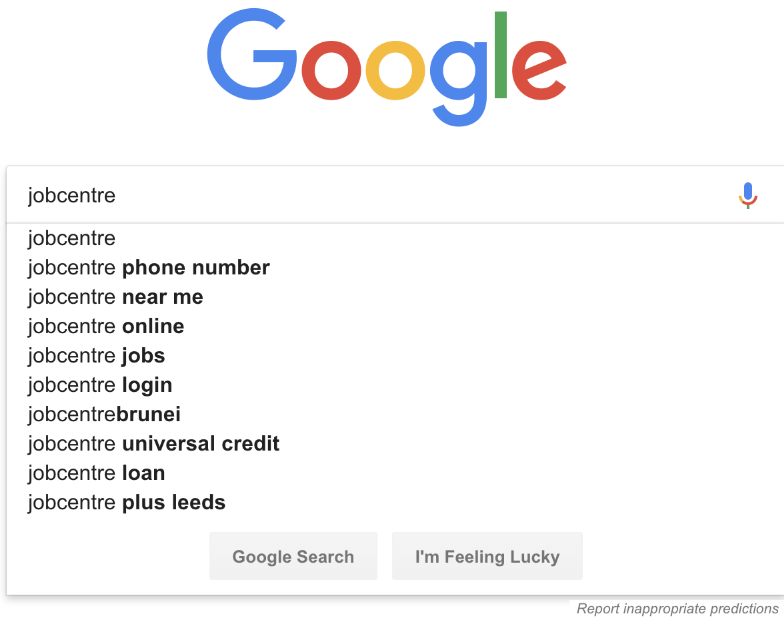


Figure 1.1: Google Autocomplete with Search Keyword: Jobcentre

1.4 Nowcasting Model for Real-time Google Search Data Selection

In this section, I present a class of penalized models for the nowcasting framework. This framework is useful in an environment in which the underlying structure of data is high dimensional with covariates that are highly correlated. In Section 1.4.2, I walk through the details of the adaptive elastic-net model and of constructing the out-of-sample forecasts. Section 1.4.3 discusses the empirical strategies.

1.4.1 Variable Selection Models

The objective is to accurately predict the change between the current value of the macroeconomic variable and last month official release, Δy_t, from the set of contemporaneous predictors, $\mathbf{X}_t$. Variable selection models, in particular, are well suited for the nowcasting framework when the number of potential predictors is large, yet the number of macroeconomic

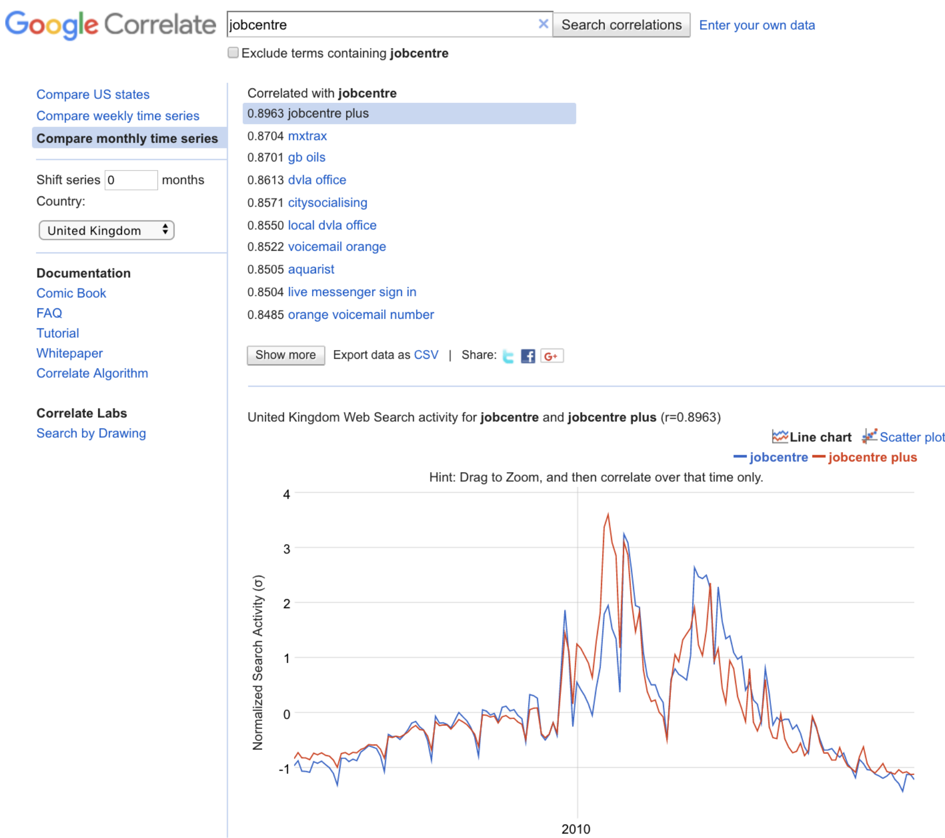


Figure 1.2: Google Correlate with Search Keyword: Jobcentre

observations is small. I am particularly interested in ensuring good prediction accuracy as well as discovering relevant predictive variables from Google search queries. Although I construct a high-dimensional set of search queries related to the macroeconomic variables, a considerable amount of these may not express predictive power. Then I assume that the true underlying nowcasting model is a sparse representation. And variable selection (Fan and Li, 2006) is particularly important addressing the sparsity problem.

When the dimension is high, an ideal variable reduction method should enjoy the oracle property (Fan and Li, 2001). That is, it selects the right subset of the model and, at the same time, its coefficient estimator achieves the optimal estimation rate. Simply put, the asymptotic distribution of the estimators should be the same as the asymptotic distribution of the maximum likelihood estimators (MLE) on only the true support of the underlying model.

To guarantee an optimal sample performance, Zou (2006) proposes an adaptive LASSO model whose adaptive weights penalize coefficient estimates with an l_1 norm. However, with only an l_1 penalization term, the model will have poor performance when there are highly correlated covariates. Obviously, Google Trends time series present this collinearity problem. To handle the multi-collinearity problem properly, Zou and Zhang (2009) introduce the adaptive elastic-net model that combines an l_2 and l_1 penalization terms and considers the adaptive LASSO shrinkage to remedy the drawbacks of the adaptive LASSO model.

In this chapter, I adopt the adaptive elastic-net model for the task of macroeconomic nowcasting with the high-dimensional Google search data. First, let us consider a naive elastic-net model (Zou and Hastie, 2005). The l_1 norm regularization term, with penalty parameter $\lambda_1 \in R_+$, is introduced to encourage the sparsity of the model. The l_2 norm regularization term, with penalty parameter $\lambda_2 \in R_+$, stabilizes the solution path of the minimization problem and improves prediction performance when the regressor space exhibits high correlation. With any fixed non-negative λ_1 and λ_2, the elastic-net regression coefficients are obtained by solving the following minimization problem:

$$\min_{\beta} \quad \sum_{t=1}^{T}(\Delta y_t - \mathbf{X}_t^T\beta)^2 + \lambda_2|\boldsymbol{\beta}|^2 + \lambda_1|\boldsymbol{\beta}|_1 \tag{1.1}$$

where

$$|\boldsymbol{\beta}|_1 = \sum_{j=1}^{p}|\beta_j| \quad \text{and} \quad |\boldsymbol{\beta}|^2 = \sum_{j=1}^{p}\beta_j^2.$$

Then the elastic-net coefficient estimator $\hat{\boldsymbol{\beta}}$ can be solved equivalently through

$$\hat{\boldsymbol{\beta}} = \arg\min \sum_{t=1}^{T}(\Delta y_t - \mathbf{X}_t^T\beta)^2 \tag{1.2}$$

$$s.t \quad (1-\phi)|\boldsymbol{\beta}|_1 + \phi|\boldsymbol{\beta}|^2 \leq c \quad \text{for some constant c} \tag{1.3}$$

where $\phi = \frac{\lambda_2}{\lambda_1+\lambda_2}$ and the inequality in (1.3) is called the elastic-net penalty[6]. When the regressors $\mathbf{X}$ are orthogonal designed, ϕ will be 0 and the minimization problem in (1.1) will be reduced to a LASSO regression. In the high-dimensional and high-correlated Google Trends data, the elastic-net penalty ($0 < \phi < 1$) can both improve LASSO prediction by encouraging group effects and stabilizing the solution path. Therefore, the combination of adaptive weights of the adaptive LASSO model proposed by Zou (2006) and the elastic-net penalty in the naive elastic-net model can improve prediction accuracy ywofold. First, the adaptive l_1 penalty helps achieve the oracle property. Second, considering an elastic-net penalty to handle collinearity in the high-dimensional setting can remedy the limitation of LASSO, which fails to do group selection.

I construct the adaptive elastic-net model in two steps. First, I need to construct the adaptive weights by solving the coefficient estimator $\hat{\boldsymbol{\beta}}_{elanet}$ from the naive elastic-net model. I set the adaptive weights as

$$\hat{\omega}_j = (|\hat{\beta}_{j,elanet})|)^{-\gamma}, \quad j = 1, ..., p \tag{1.4}$$

where γ is some positive constant. Next, the adaptive elastic-net model coefficient estimators are obtained via

[6]The elastic-net penalty is the convex combination of the l_1 and l_2 penalties.

$$\hat{\beta}_{Aelanet} = \arg\min \sum_{t=1}^{T} (\Delta y_t - \mathbf{X}_t^T \beta)^2 + \lambda_2 |\boldsymbol{\beta}|^2 + \lambda_1' \sum_{j=1}^{p} \hat{\omega}_j |\beta_j|. \tag{1.5}$$

Let $\mathcal{A} = \{j : \hat{\beta}_{j,Aelanet} \neq 0\}$, and p is the size of $\mathcal{A}$. The adaptive elastic-net model enjoys the oracle properties. Argued in Fan and Li (2001), a desirable variable selection framework should follow an oracle procedure: (1) identifies the right subset model: $\{j : \hat{\beta}_j \neq 0\} = \mathcal{B}$; and (2) has asymptotic normality, $\sqrt{n}(\hat{\beta}(\lambda)_{\mathcal{B}} - \beta^*_{\mathcal{B}}) \to_d N(0, \Sigma^*)$, where Σ^* is the covariance matrix knowing the true subset model.

Some papers have been concerned with the choice of penalty parameters λ. Appropriate λ will lead to a consistent LASSO estimator with the $(s\log p/n)^{1/2}$ [7] rate of convergence. Bickel et al. (2009) propose that the penalty estimator be chosen following the Bickel-Ritov-Tsybakov rule[8]. However, in practice, it is commonly and recommended that cross-validation be used to choose the optimal λ. Chetverikov, Liao, and Chernozhukov (2017) show that the cross-validated LASSO estimator achieves the fastest possible rate of convergence when the penalty parameter λ for the estimator is chosen using K-fold cross-validation. For the adaptive elastic-net model, I will employ the K-fold cross-validation procedure to select proper penalty parameters λ_1' and λ_2. K-fold cross-validation divides the data set randomly into K different subsets, and typically I pick K = 5 or 10 in practice. Then the model is trained and estimated over (K-1) sets for a range of values of the parameter λ, leaving one of the subsets as the validation set. The process is repeated by using each of the K subsets as a validation set, and it yields K estimates of the mean squared error for each parameter value. Finally, the K-fold estimate is the average of the set of K estimates. Mathematically, let Ω_n be a set of candidate values of λ. For $k = 1, ..., K$, $\lambda \in \Omega_n$, and I_k is the validation set. Let

$$\hat{\beta}_{-k}(\phi') = \underset{\beta \in R_p}{\operatorname{argmin}} \left(\frac{1}{n - n_k} \sum_{t \notin I_k} (\Delta y_i - \mathbf{X}_t^T \beta)^2 + \phi' |\boldsymbol{\beta}|^2 + (1 - \phi')' \sum_{j=1}^{p} \hat{\omega}_j |\beta_j| \right) \tag{1.6}$$

[7] $s = \sum_{i=1}^{p} 1\beta_i \neq 0$.

[8] $\lambda = 2c\sigma n^{-\frac{1}{2}} \Phi^{-1}(1 - \alpha/(2p))$, where $c > 1$ and $\alpha \in (0, 1)$ are constants, σ is the standard deviation of ϵ, and σ is typically estimated from the data.

be the adaptive elastic-net estmator corresponding to all observations excluding the ones in the validation set k and $\phi' = \frac{\lambda_2}{\lambda_1' + \lambda_2}$. Then the cross-validation choice of ϕ' is

$$\hat{\phi}' = \underset{\lambda \in \Omega_n}{\text{argmin}} \quad CV(\lambda) = \frac{1}{n} \sum_{k=1}^{K} \sum_{t \notin I_k} (\Delta y_t - \mathbf{X}_t^T \hat{\beta}_{-k}(\phi'))^2. \tag{1.7}$$

The best penalty parameter value of ϕ' is the one with the lowest K-fold estimate. It is also typical to report the minimum parameter value such that its K-fold estimate does not exceed the minimum K-fold estimate by more than one standard error using a lower number of covariates to achieve a parsimonious model. Simply put, I can take the "most penalized" model whose error is within one standard error of the minimal error. Recall that I can compute standard errors for the cross-validation error curve at each penalty parameter ϕ'. For each cross-validation,

$$CV_k(\phi') = \frac{1}{n_k} \sum_{t \in I_k} (\Delta y_t - \mathbf{X}_t^T \hat{\beta}_{-k}(\phi')^2 \tag{1.8}$$

where n_k is the number of points in the k_{th} fold. Therefore, the sample standard error of $CV_1(\phi'), ..., CV_k(\phi')$ is

$$SE(\phi') = \frac{\sqrt{var(CV_1(\phi'), ..., CV_K(\phi'))}}{\sqrt{K}}. \tag{1.9}$$

In practice, the one-standard-error rule is an alternative way of choosing ϕ' from the cross-validation curve. Start with the usual estimate

$$\hat{\phi}' = \underset{\phi' \in \Omega_n}{\text{argmin}} \quad CV(\phi') = \frac{1}{n} \sum_{k=1}^{K} \sum_{i \notin I_k} (\Delta y_i - \mathbf{X}_t^T \hat{\beta}_{-k}(\phi'))^2 \tag{1.10}$$

and move ϕ' in the direction of increasing penalty until it reaches

$$CV(\phi') \leq CV(\hat{\phi}') + SE(\hat{\phi}'). \tag{1.11}$$

Recall that a desirable variable selection model should identify or recover the true model. Choosing the optimal $\hat{\phi}'$ with the K-fold cross-validation method that achieves the smallest cross-validation error often corresponds to a large model (not too much penalization). Then applying the one-standard-error rule, which increases the penalty parameter, will be helpful in recovering the true model.

In an ultra-high-dimensional setting in which the number of predictors is larger than the number of observations, Zou and Zhang (2009), and Fan and Lv (2008) suggest using the Sure Independence Screening method first to reduce the high dimensionality to a smaller scale ($d < n$), then applying the adaptive elastic-net model to the selected predictors. Since I will have more than 300 predictors that are relevant search queries from Google Trends, I consider using Sure Independence Screening and then conduct the nowcasting task with the adaptive elastic-net model.

1.4.2 Variable Selection Model Application in Macroeconomic Nowcasting

In this section, I use the adaptive elastic-net models to construct nowcasts for unemployment rate or inflation movements for different countries. For each dependent variable, I have more than 300 potential predictors and expect a parsimonious model that contains true underlying predictors. Therefore, the task is to discover these important search queries and obtaining an accurate prediction.

Computation of the adaptive elastic-net solutions is a quadratic programming problem that can be tackled using standard numerical analysis algorithms. Efron et al. (2004) propose the least angle regression (LARS), procedure which exploits the special structure of the minimization problem, and provides an efficient way to compute solutions simultaneously for all values of λ. First, I need to standardize the predictors to have mean zero and unit norm. The algorithm starts with all coefficients $\beta_1 = \beta_2 = ... = \beta_p$ equal to zero. Second, it finds the predictor x_i most correlated with the residual r, $y - \bar{y}$. Then it increases the coefficient β_i in the direction of the sign of the correlation of the residual and x_i until some other regressor x_j has as much as correlation with the residual r as x_i, and calculate the new residual of y and fitted $\hat{y}$. Fourth, move the two coefficients β_i and β_j in the least squares direction for x_i and x_j until some other regressor x_m has as much as correlation with the current new residual. Finally, it continues the procedure until all predictors are included in the model. It stops when correlation of the current residual and the rest of the regressors is zero. The adaptive elastic-net regression can be efficiently computed with the

LARS algorithm[9].

1.4.3 Empirical Strategy

I obtain real-time monthly macroeconomic data from 2004:M1 to 2019:M1 for the United States, the United Kingdon, Japan, and Australia. All of the macroeconomic data come from the OECD Original Release and Revision Database. In particular, I use the personal consumption expenditure index to measure inflation (PCE) for the U.S., the harmonized consumer price index (HICP) for the U.K., and core consumer prices (CPI) for Japan and Australia. Unemployment rates are from the OECD Original Release and Revision Database.

As for the nowcasting model, I consider the framework in Section 1.4:

$$\Delta y_t^c = \gamma \Delta y_{t-1}^c + \boldsymbol{X_t^c}' \beta_1 + \epsilon_t \tag{1.12}$$

where Δy_t^c is the change in the unemployment rate or inflation rate of country c and $\boldsymbol{X_t^c}$ is high-dimensional Google search data corresponding to the dependent macroeconomic variable y_t^c. The model also includes lags of the dependent variable up to 2 and an intercept. Since Google Trends search data present strong monthly seasonality, I use the additive Holt-Winters exponential smoothing method to extract out seasonal patterns. Also, each regressor has been standardized with mean zero and unit norm before fitting in the adaptive elastic-net model. For the sake of notation simplicity, I do not explicitly indicate these lagged terms in (1.12). Also, I consider an AR model and a "no change" random walk model as the benchmarks.

Then I carry out recursive regressions to obtain out-of-sample forecasts from 2009:M4[10] to 2019:M1. I assume the following timing convention. At the end of month $(t+1)$ or early in month $(t+2)$, The macroeconomic variables such as unemployment rate and information have not been observed or officially released. When the Google search data for the last week of month $(t+1)$ becomes available, the out-of-sample nowcasts of the depend variable Δy_{t+1}

[9]Both predictors and regression coefficients are adjusted by the adaptive weights in the algorithm.

[10]The first forecast is at 2009:M4, because I consider the first-differenced series due to stationary issues and also include the first two lags of the dependent variable.

can be obtained as follows:

$$\Delta \hat{y}_{t+1} = \boldsymbol{X_{t+1}}^T \beta(\lambda). \tag{1.13}$$

I establish a three-step procedure for variable selection with the adaptive elastic-net model. The third step is conducting an OLS post-adaptive elastic-net estimation that only includes covariates with non zero coefficients of the adaptive elastic-net λ-1se [11] specification. This procedure ensures at most the same number of non zero coefficients as the adaptive elastic-net model. Weights are used to penalize different coefficients to include the oracle properties, which are related to identifying the right subset model and having the optimal estimation rate. In the case with Google search data, only a small number of covariates are the true factors correlated with the outcome, despite introducing a large number of covariates at the initial stage. Since there exist groups of variables among which pairwise correlations are very high in some regressors, the model encourages group selection as well.

When the number of regressors is much larger than the number of observations, this will create challenges to scalable statistical inference and computational efficiency. In that case, I apply the Sure Independence Screening method to help reduce the dimensionality of the regressor space from a very large scale to a moderate size. Applying marginal regressions, the SIS procedure ranks all regressor candidates based on their marginal correlation with the dependent variable and keeps the top d_m regressors. Namely, the set of regressors, $\hat{\mathcal{S}}$, retained by the SIS is defined as

$$\hat{\mathcal{S}} = \{|corr(x_i, y)| \text{ is the top } d_m \text{ largest correlation}\},$$

where $1 \leq d_m \leq p$ and p is the number of regressor candidates. In the case of my model, I choose $d_m = n - 1$[12], and n is the size of the recursive window for each forecasting iteration. Then the simple elastic-net regression is performed to obtain an initial estimator of the coefficients and serves as the weighting parameters of the adaptive elastic-net in the second

[11] λ one standard deviation rule which is discussed in the previous section.

[12] Fan and Lv (2008) they consider $d_m = n - 1$ or $d = \frac{n}{log(n)}$.

stage. The optimal regularization parameter λ is selected according to the five fold cross-validation. For the third stage, I fit an OLS regression with selected independent variables from the adaptive elastic-net model to obtain consistent predictors.

I use the first 60 months to estimate the historical relationship between the dependent variable and relevant Google search data. The first recursive one-month-ahead forecast starts from 2009M4. Intuitively, these out-of-sample forecasts are illustrated in Figure 1.5-Figure 1.16, which graph actual and out-of-sample nowcasted changes using three different models with or without inclusion of Google search data. Forecasts from the adaptive elastic-net model track actual changes in inflation in the U.K., Japan, and Australia very well. For example, as shown in Figure 1.5, the largest monthly change in inflation in the U.K. during 2009M4 to 2019M1 took place in 2009M12, when the inflation rate increased by 0.60 %. The out-of-sample forecast of the adaptive elastic-net model signals an increase by 0.72 percent, while the AR model only predicts a less than 0.30 percent increase one month behind. The reason is that extrapolated past information from inflation alone adds minimal predictive power to the current change. Similarly, in Figure 1.20 I observe that the adaptive elastic-net forecasts follow the course of unemployment changes very well, especially during the periods from 2009 to 2014 in the case of the U.S.

1.5 Empirical Results

1.5.1 Evaluating Unconditional Point Forecasts

To conduct a rigorous evalualation of the nowcasting performance of the adaptive elastic-net model vis-a-vis benchmarks such as the autoregressive and random walk models, I consider comparing mean squared errors (MSE) and use the DMW test. In particular, the DMW test is used to assess whether forecasts from the model that includes Google Trends data is significantly different from forecasts from the benchmark models. I also consider the Newey-West long-run variance estimator to control for auto correlation in the DMW test.

Panel A of Table 1.3 presents the ratio of out-of-sample nowcasts of the adaptive elastic-net model with inclusion of Google search data to those of the autoregressive model for both unemployment rate and inflation changes in four countries. The ratios of MSEs are smaller than one for all macroeconomic variables except for inflation in Japan and the U.S.. This implies that the inclusion of Google data improves nowcasting performance compared to the benchmark models. However, a simple autoregressive model outperforms the nowcasting model in the case of U.S. inflation. As Atkeson and Ohanian (2001) state, "none of the forecasts is more accurate than the naive forecast." And as described by Janet Yellen (2017), the former Chair of the Board of Governors for the U.S. Federal Reserve, "our framework for understanding inflation dynamics could be mis-specified in some fundamental way." In the case of Japan, although the mean squared prediction error of the autoregressive model is smaller than the adaptive elastic-net specification, the magnitude of each point forecast is centered around zero. Moreover, most of the point forecasts signal the opposite direction to the true realized inflation change. Based on the DMW statistic and p-value, the performance of the nowcasts does not differ much from the autoregressive model for most of the macroeconomic variables.

Panel B of Table 1.3 presents the ratio of out-of-sample nowcasts of the adaptive elastic-net model with inclusion of Google search data to those of the random walk model for both unemployment rate and inflation changes in four countries. The ratios of MSEs are all smaller than one for all macroeconomic variables. Table 1.3 also contains the DMW test statistics and their p-values. Almost all of my nowcasts reject the null hypothesis at the 1% significance level based on the DMW test statistic. Still, I fail to conclude that the nowcasts of U.S. inflation outperform the naive benchmark.

In terms of dimensionality reduction, the adaptive elastic-net with the λ-1se specification is able to select on average around 37 predictors for the U.S. inflation rate from 108 estimation windows, 32 for the U.S. employment rate, 27 for the U.K. inflation, 31 for the U.K. unemployment rate. For the robustness check, I also consider the Clark and West (CW) test, as well as different forecasting horizons. The results are included in Appendix 1.7. I conclude that my point forecasts with inclusion of Google search data demonstrate the greatest predictive power for nowcasting these macroeconomic variables.

Table 1.3: Point Nowcast Test Result

	Panel A: DMW Test against AR Model			
	Δ Unemployment		Δ Inflation	
	MSE Ratio	DMW Test	MSE Ratio	DMW Test
U.S.	0.908	0.807	1.207	−1.016
U.K.	0.985	0.111	0.781	2.593 * **
Japan	1.011	−0.087	1.139	-0.841
Australia	0.698	2.794 * **	0.966	0.373
	Panel B: DMW Test against Random Walk Model			
U.S.	0.527	3.573 * **	0.902	0.544
U.K.	0.219	4.157 * **	0.516	3.525 * **
Japan	0.440	4.732 * **	0.605	2.091 * *
Australia	0.309	5.653 * **	0.559	3.706 * **

Note: The success ratio is computed as the number of correct nowcasts divided by the total number of out-of-sample nowcasts. Here, I have 108 out-of-sample nowcasts for each variable. The table also reports the t-statistics from the weighted directional test for each variable. The test uses the Newey-West LRV estimator, which controls for auto correlation. ***, **, and * represent the 1%, 5%, and 10% significance level, respectively.

1.5.2 Point Forecasts of Macroeconomic Variables: Conditional Superior Predictive Ability

In the previous section, I conducted an unconditional superior predictive ability test on my weighted directional forecasts. Although the DMW test is informative about average model performance against the benchmark, I notice that the inclusion of Google search data is helpful for signaling big movements in the data. I need a statistical test to evaluate the relative performance of different models over a business cycle. Specifically, I are interested in testing *when* the out-of-sample forecasts are better than the benchmarks such as AR and random walk models. Here I consider the CSPA test proposed by Li, Liao, and Quaedvlieg (2019) to assess relative predictive performance based on a cyclical indicator variable (e.g., average inflation over the past months), which tracks a country's business cycle.

I consider the following squared error loss function difference $(Z_t)_{1\leq t\leq T}$ to evaluate the performance of the benchmark model relative to my forecast method:

$$Z_t = (Y_t - Y_t^*)^2 - (Y_t - Y_{0,t}^*)^2, \tag{1.14}$$

where $(Y_t^*)_{1\leq t\leq T}$ are the out-of-sample forecasts from the adaptive elastic-net model and $(Y_{0,t}^*)_{1\leq t\leq T}$ are the forecasts from the benchmarks such as AR and random walk model. If $Z_t \leq 0$, then the competing forecasts $(Y_t^*)_{1\leq t\leq T}$ outperform the benchmart at time t. Extending the influential paper by Giacomini and White (2006) about conditional equal predictive ability, Li, Liao, and Quaedvlieg (2019) propose a conditional superior predictive ability (CSPA) hypothesis. The null hypothesis states that the benchmark model weakly outperforms the competing model conditional on the information set S_t, which describes the business cycle of an economy. Rejecting this null hypothesis means that the competing forecast method is weakly dominant over the benchmark model. The null hypothesis of the CSPA can be written as

$$H_0 : E(Z_t|S_t = s_t) \geq 0 \quad \text{almost surely, } t = 1, 2, ... \tag{1.15}$$

Testing the conditional moment inequality requires that I estimate conditional expected loss differences using functional inference. Using the intersection-bound method (Chernozhukov, Lee, and Rosen, 2013), the null hypothesis can be written as

$$H_0 : \eta \equiv \inf_{s\in\mathcal{F}} E(Z_t|S_t = s_t) \geq 0. \tag{1.16}$$

The $(1-\alpha)$ upper confidence bound can be constructed as follows:

$$\lim_{n\to\infty} \inf \mathrm{P}(\eta \leq \hat{\eta}_{1-\alpha}) \geq 1 - \alpha. \tag{1.17}$$

The alternative test rejects the null when $\hat{\eta}_{1-\alpha} < 0$, with type I error bounded by α in a sufficiently large sample.

Constructing the CSPA test involves estimating the conditional expected loss differential function non-parametrically with least squares regressions. Details of the construction can be found in Li, Liao, and Quaedvlieg (2019). In this empirical exercise, I consider the moving average of inflation as the conditional variable. Then I estimate the expected mean squared error differential function non-parametrically with the polynomial series as approximating basis functions. The number of polynomials is chosen with minimum Akaike Information Criteria (AIC) by using five fold cross-validation. Appendix 1.8.3 lays out the details of the implementation of the CSPA test.

Table 1.4 reports the results of the CSPA test. It gives the value of $(1-\alpha)$ upper confidence bound for the null hypothesis of the CSPA test. Following Li, Liao, and Quaedvlieg (2019), the functional inference relies on intersection bounds that are defined by the infimum of a nonparametric function. The second and fourth columns report the values of the 99% upper confidence bound for the null hypothesis of the CSPA test. They clearly show that $\hat{\eta}_{1-\alpha}$ for all of the fundamental variables I am interested in are below zero. Table 1.4 shows strikingly different results from Table 1.3. These results indicate that all forecasts of the adaptive elastic-net model are rejected with respect to the benchmark AR model in terms of the conditional forecasts. In the unconditional setting, according to the DMW test results, only the forecasts for the unemployment rate in Australia and inflation in U.K. dominate the AR model. Therefore, the result of the CSPA test give us another narrative for evaluating the forecasts during heterogeneous economic conditions. Clearly, all of the forecasts perform better than the benchmarks for at least some periods during the time from 2009M4 to 2019M1.

In the case of Japan, while the unconditional average mean squared error of the adaptive elastic-net model is slightly higher than that of the benchmark AR model, Figure 1.33(a) shows that the forecasts based on Google search data obtain improved forecast performance during the period of high inflation, yet with a wide confidence bound. Also, the figure suggests the relatedness of the confidence bound to the distribution of inflation in Japan. During protracted periods of deflation after the crisis, the Bank of Japan kept eliminating cyclical slack and injecting a combination of asset purchases and yield management to lift Japan out of the deflation trap. However, the collective wisdom of people's search behavior provides only insight for nowcasting changes in inflation during the protracted periods.

In contrast, in the case of the U.S., point forecasts from the AR model outperform the

Table 1.4: Conditional Superior Prediction Ability (CSPA) Test Results

	Δ Unemployment		Δ Inflation	
	$\hat{\eta}_{1-\alpha}$	Rejection	$\hat{\eta}_{1-\alpha}$	Rejection
Panel A: CSPA Test Rejection Result against the Benchmark AR Model				
U.S.	-0.296	Yes	-0.231	Yes
U.K.	-0.015	Yes	-0.028	Yes
Japan	-0.030	Yes	-0.124	Yes
Australia	-0.025	Yes	-0.062	Yes
Panel B: CSPA Test Rejection Results against the Benchmark Random Walk Model				
U.S.	-0.561	Yes	-0.336	Yes
U.K.	-0.002	Yes	-0.044	Yes
Japan	-0.087	Yes	-0.404	Yes
Australia	-0.084	Yes	-0.075	Yes

Note: This table shows the value of the 99% upper confidence bound for the null hypothesis of the CSPA test. If $\hat{\eta}_{99\%} < 0$, then reject the CSPA hypothesis that the benchmark model weakly outperforms the alternative competing model uniformly across all periods specified by the conditioning variable s_t.

adaptive elastic-net on average. Figure 1.29(a) clearly shows that the nowcasts perform exceptionally well during negative inflation periods. This implies that the adaptive elastic-net model is preferred during the crises. In addition, the conditional performance of the model shows prediction power for the unemployment rate most of the time with a very tight confidence bound. Compared to the DMW test, the CSPA test provides richer insights or information for predicting changes in macroeconomic variables of the United States.

Figure 1.31 shows that the point forecasts outperform two benchmarks at the majority of states of the conditioning variable, especially at the first quantile of inflation distribution (low inflation) in the U.K. The result also coincides with the DMW test in Table 1.3.

Our conditional point forecasts also reject the null hypothesis of the CSPA in the case of Australia. In Figure 1.35 and Figure 1.36, I observe that although the expected conditional forecast function is never significantly smaller than zero, it is typically negative for a large part of the range of conditioning variables.

1.6 Forecasting Interest Rate Differentials and a Simple Strategy foExchange Rate Forecasting

In this section, I demonstrate the directional forecastability of interest rate differentials using Google search data. I also construct a simple exchange rate forecasting strategy using "surprises" from interest rate differential forecasts.

When the Fed and other central banks set their short-term nominal interest rate according to the Taylor rule, following Taylor (1993), the monetary policy rule is formulated as

$$i_t = \pi + \pi_t + \phi(\pi - \bar{\pi}) + \gamma y_t + r, \tag{1.18}$$

where π_t is the inflation rate, and y_t is the output gap (the percent deviation of actual real GDP from an estimate of its potential level), and r is the real federal funds rate (usually 2%). Sometimes I can replace the GDP gap with the unemployment gap in specification (1.18). In the previous section, I successfully showed the nowcastability of the unemployment rate and inflation using Google search data. This implies that these nowcasts can be employed as proxies for the current state of labor markets and prices. Therefore, once I obtain the nowcasts before the release of official government data, I hypothesize that I can further forecast the short-term nominal interest rate using the same set of relevant Google search covariates. Specifically, I consider the following interest rate differential forecasting equation:

$$\begin{aligned}\Delta i_{t+1} - \Delta i^*_{t+1} = \alpha &+ \beta_1 \Delta \text{unemployment}^{Google}_t + \beta_2 \Delta \text{inflation}^{Google}_t \\ &+ \beta_3 \Delta \text{unemployment}_{t-1} + \beta_4 \Delta \text{inflation}_{t-1} \\ &+ \gamma_1 \Delta \text{unemployment}^{Google*}_t + \gamma_2 \Delta \text{inflation}^{Google*}_t \\ &+ \gamma_3 \Delta \text{unemployment}^*_{t-1} + \gamma_4 \Delta \text{inflation}^*_{t-1} + \epsilon_t,\end{aligned} \tag{1.19}$$

where $(\cdot)^{Google}$ is the covariate space from Google search data of the U.S. and $(\cdot)^{Google*}$ for the foreign counterpart. Every month, I fit an adaptive elastic-net model with relevant Google search queries. Particularly, I consider the rolling window forecasting scheme with a fixed window size of 60 month and obtain one-month-ahead forecasts of changes in the interest rate differential, $\hat{ID}_t$. The relationship between short-term interest rates and exchange rate is important for policy makers and investors. This section aims to exploit the predictability of surprises in short-term interest rates to improve the forecasts of exchange rate movements. I follow a simple forecasting strategy for the exchange rate movements based on the surprises in my interest rate differential forecasts. A surprise takes place when the sign of the interest rate differential point forecast switches after two consecutive periods. Namely, I forecast appreciation (depreciation) in the U.S. dollars between t and $t+h$ ($h = 1, 3, 6, 9,$ and 12) if a surprise of increase (decrease) in the interest rate differential takes place. Explicitly, I form the rule below to make forecasts for exchange rate movements:

$$D_{t+h} = \begin{cases} 1, & \text{if} \quad \hat{ID}_t > 0, \hat{ID}_{t-1} < 0, \hat{ID}_{t-2} < 0 \\ -1, & \text{if} \quad \hat{ID}_t < 0, \hat{ID}_{t-1} > 0, \hat{ID}_{t-2} > 0 \\ 0, & \text{otherwise.} \end{cases}$$

I illustrate this forecasting strategy in Figure 1.3 and Figure 1.4. Figure 1.3 displays the surprises from the interest rate differential. I plot these surprises in the interest rate differential and mark their performance with different colors and sizes. I place the big orange circle on time $(t+1)$'s interest rate differential if the surprise of increase for $(t+1)$

is made successfully at time t, small dark brown circle if it is incorrectly. Similarly, if the surprise of decrease forecast is made successfully at time t, I mark a big green circle on the time $(t+1)$, and a small dark green circle otherwise.

Figure 1.4 shows surprise-based directional forecasts for the U.S./U.K. exchange rate. Explicitly, based on the surprises from the interest rate differential forecasts, I make 6-month-ahead directional forecasts for U.S./U.K. movements. I plot the U.S./U.K. exchange rate as well as the forecasting performance with different colors and sizes. Again, a big orange circle is placed at time $(t+1)$ when the appreciation forecast for $(t+1)$ is made successfully at time t, and a small dark orange circle if it is incorrect. Similarly, if the depreciation forecast for $(t+1)$ is made successfully at time t, I mark a big green circle on the time $(t+1)$'s exchange rate. and a small dark green circle otherwise.

Intuitively in Figure 1.4, 18 out of 30 markers are either big green circles or big orange circles in Figure 1.4. This generates the 60% success ratio of U.S./U.K. exchange rate movements. Notice that although I do not make forecasts every month, the forecasts tend to capture big moves in currency during the periods I consider. I report the success ratios for three currency pairs, as well as different forecasting horizons, in Table 1.5. The success ratio is defined as the number of successful appreciation or depreciation forecasts divided by the total number of depreciation and appreciation forecasts. For the British Pound, the forecast accuracy achieves 60% at a 6-month horizon. In contrast, one-month-ahead forecasts for the Australian dollar achieve the highest success ratio over other forecasting horizons. For the Japanese yen, the success ratio is greater at longer forecasting horizons.

Figure 1.3: Performance of One-month-ahead Directional Forecasts for the Interest Rate Differential (U.K.-U.S.)

Figure 1.4: Performance of One-month-ahead Directional Forecasts for the U.S./U.K.

Table 1.5: Directional Forecast Success Ratios

Currency	1 month	3 month	6 month	9 month	12 month
U.S./U.K.	0.45	0.50	0.60	0.45	0.50
Japan/U.S.	0.61	0.44	0.46	0.52	0.62
U.S./Australia	0.55	0.48	0.45	0.45	0.40

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts. Out-of-sample forecasts start from 2009M4 to 2019M1.

1.7 Conclusion

In Chapter 1, I have found that variable selection models such as the adaptive elastic-net model which includes relevant Google Trends search data, tends to outperform naive models such as autoregressive and random walk in nowcasting or forecasting short-term macroeconomic variables.

The following main conclusions emerge: First, this paper contributes to the nowcasting literature by examining real-time Google search data, and yields improvements in nowcasting macroeconomic aggregates for four countries. Rather than remaining silent about constructing the predictors, I provide a transparent but comprehensive procedure to include Google variables in the nowcasting model. Also, the steps I take come very close to anwsering the question "what are people searching online when they face unemployment or worry about prices?"

Second, I adopt the adaptive elastic-net model in a recursive forecasting scheme to select the most important predictors in a rich data environment. Based on the DMW test, seven out of eight macroeconomic nowcasts reject the null hypothesise when the benchmark is random walk. The U.S. inflation nowcasts fail to reject the null hypothesis nevertheless. Only inflation in the U.K. and the unemployment rate in Australia significantly outperform the autoregressive model.

Third, I further find that Google search variables perform heterogeneously during expansion periods from recessions. According the CSPA test, I actually see improved performance of my nowcasts, especially in low inflation periods.

Figures

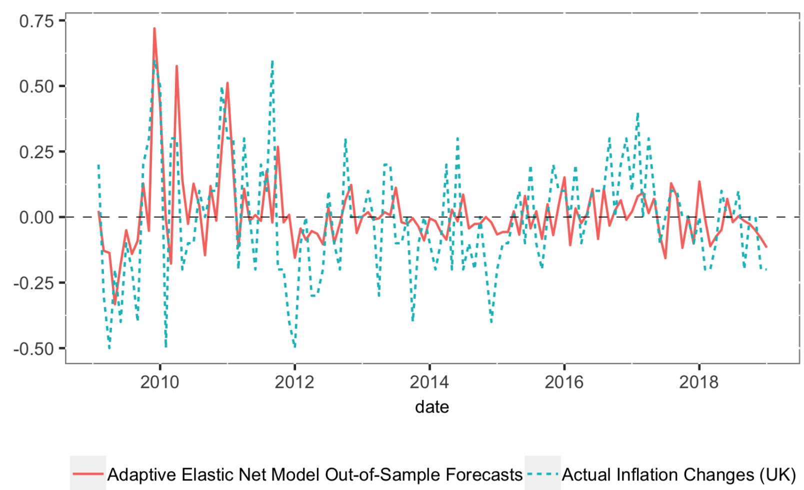


Figure 1.5: Actual and Predicted (Adaptive Elastic-net Model) Changes in Inflation (U.K.)

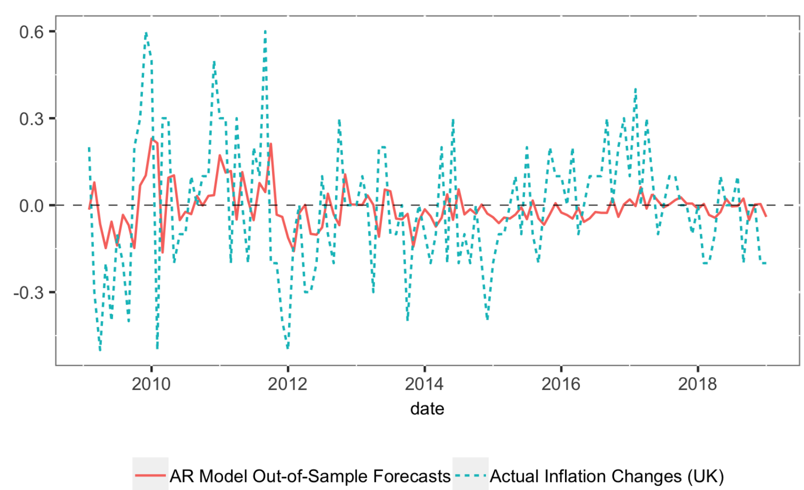


Figure 1.6: Actual and Predicted (Autoregressive Model) Changes in Inflation (U.K.)

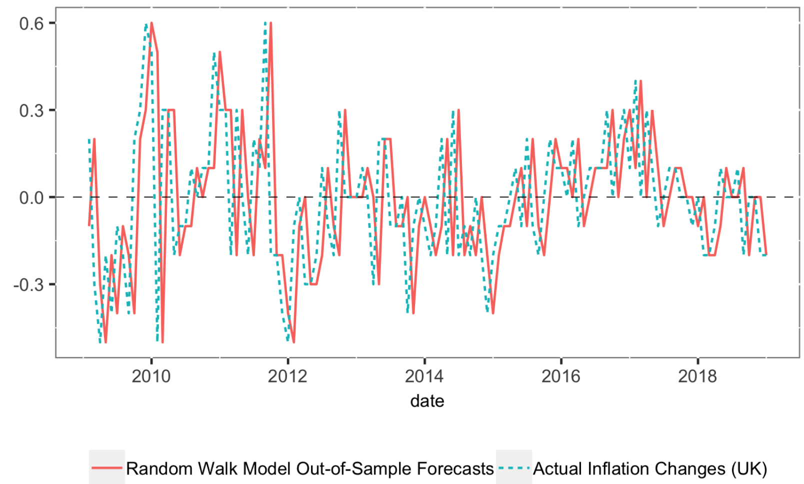


Figure 1.7: Actual and Predicted (Random Walk Model) Changes in Inflation (U.K.)

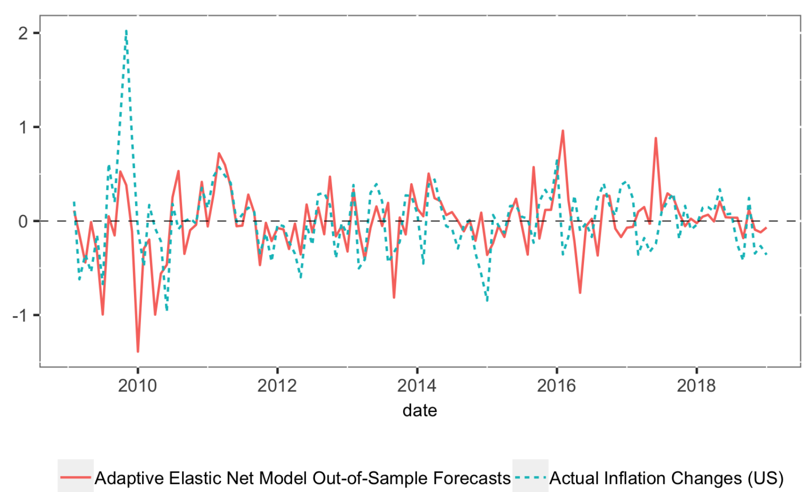


Figure 1.8: Actual and Predicted (Adaptive Elastic-net Model) Changes in Inflation (U.S.)

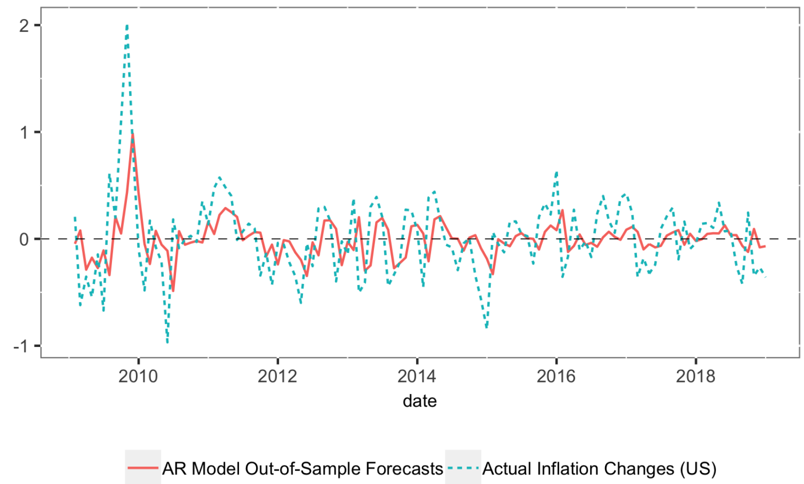


Figure 1.9: Actual and Predicted (Autoregressive Model) Changes in Inflation (U.S.)

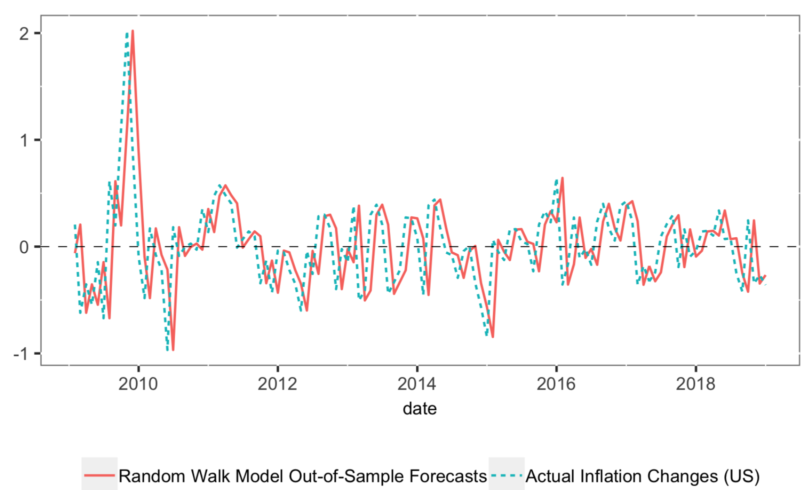


Figure 1.10: Actual and Predicted (Random Walk Model) Changes in Inflation (U.S.)

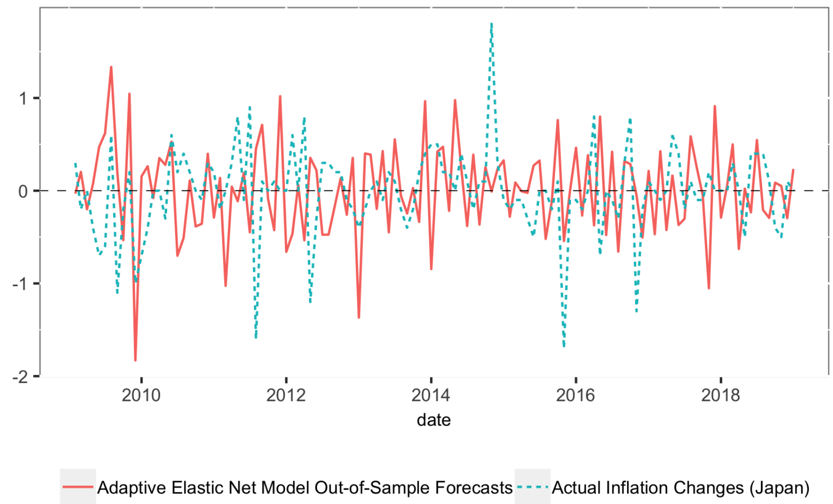


Figure 1.11: Actual and Predicted (Adaptive Elastic-net Model) Changes in Inflation (Japan)

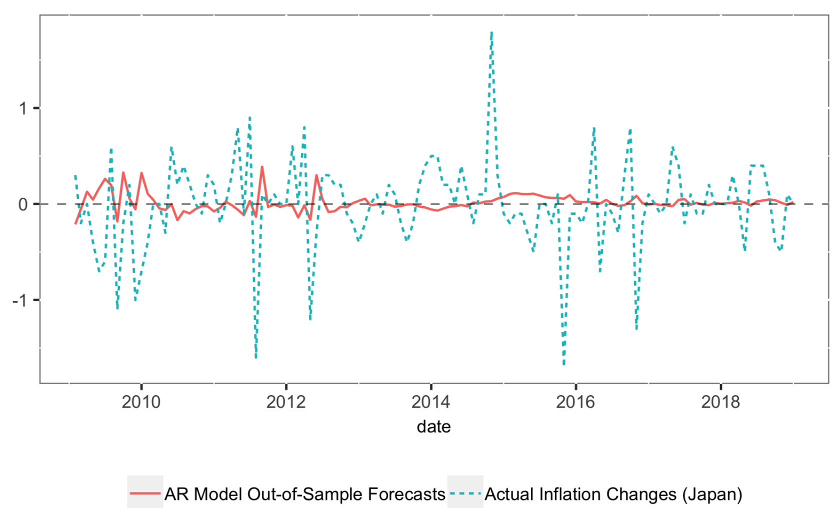


Figure 1.12: Actual and Predicted (Autoregressive Model) Changes in Inflation (Japan)

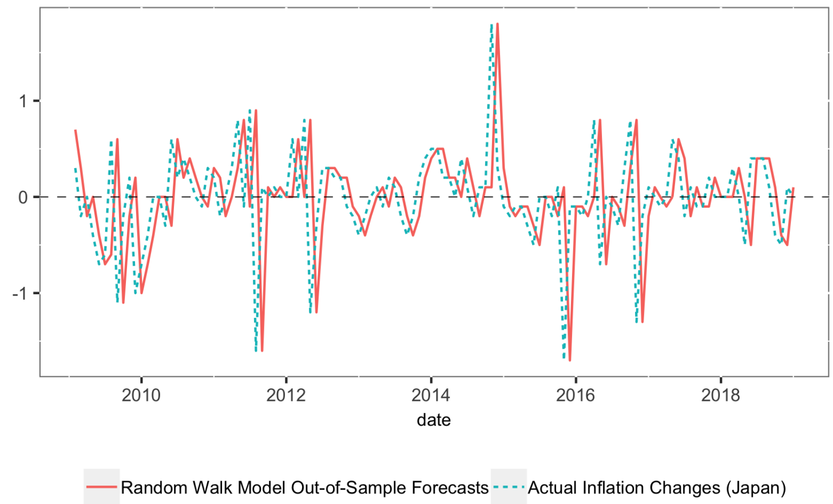


Figure 1.13: Actual and Predicted (Random Walk Model) Changes in Inflation (Japan)

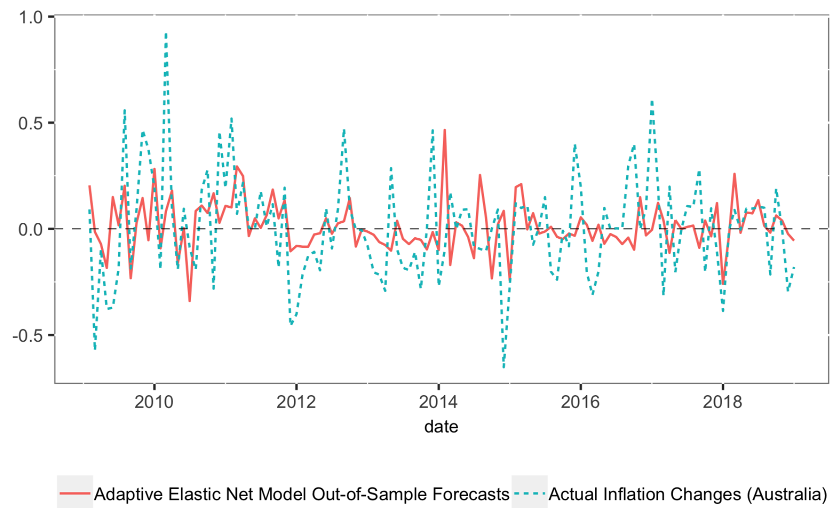


Figure 1.14: Actual and Predicted (Adaptive Elastic-net Model) Changes in Inflation (Australia)

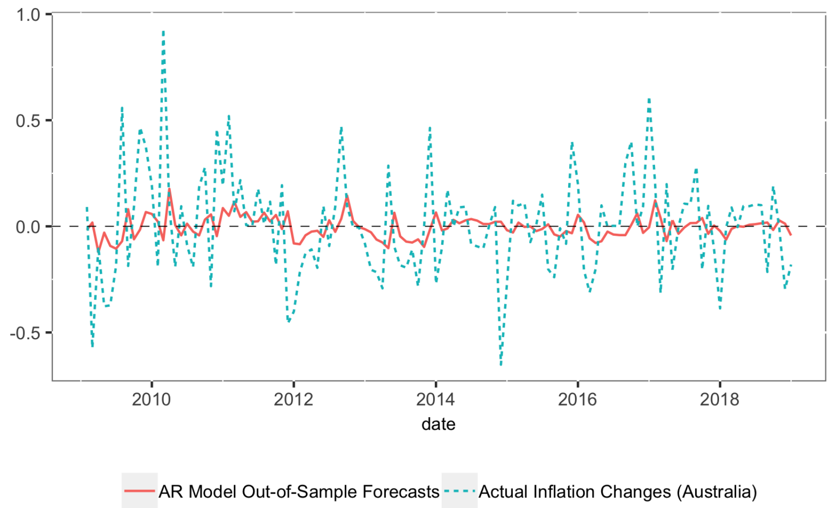


Figure 1.15: Actual and Predicted (Autoregressive Model) Changes in Inflation (Australia)

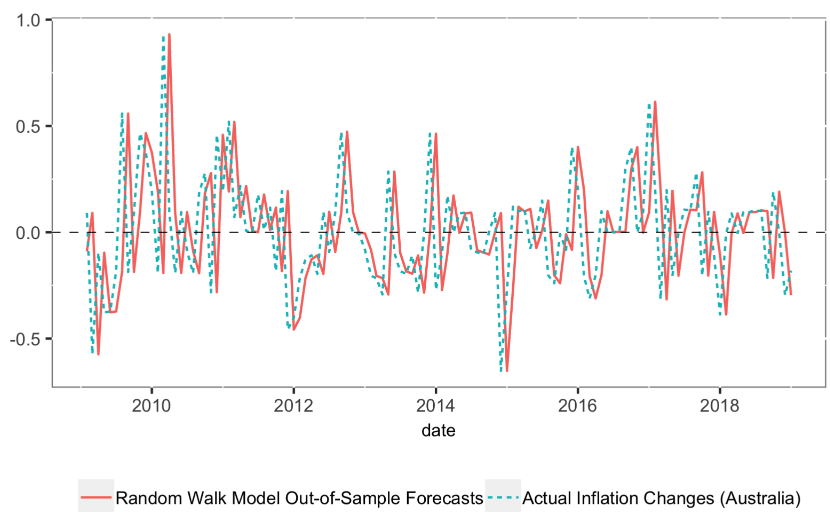


Figure 1.16: Actual and Predicted (Random Walk Model) Changes in Inflation (Australia)

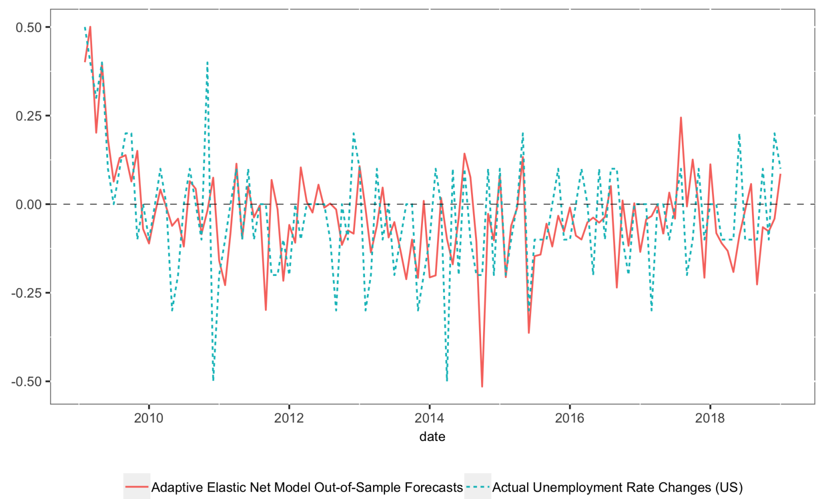


Figure 1.17: Actual and Predicted (Adaptive Elastic-net Model) Changes in the Unemployment Rate (U.S.)

Figure 1.18: Actual and Predicted (Autoregressive Model) Changes in the Unemployment Rate (U.S.)

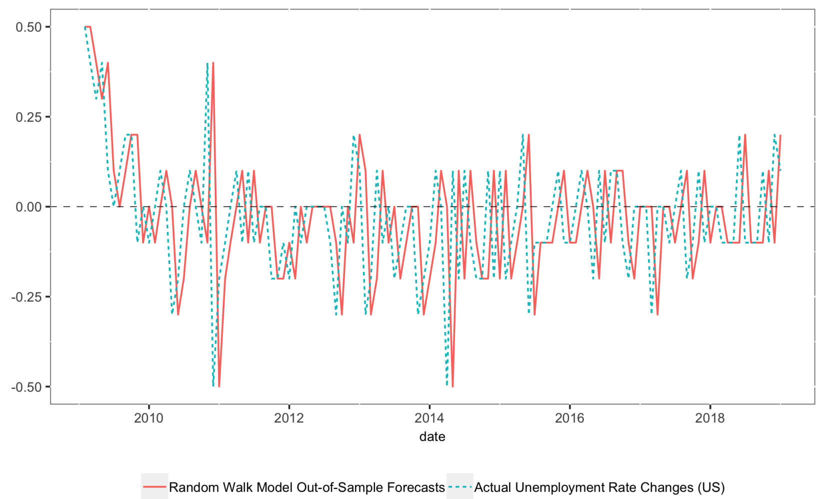


Figure 1.19: Actual and Predicted (Random Walk Model) Changes in the Unemployment Rate (U.S.)

Figure 1.20: Actual and Predicted (Adaptive Elastic-net Model) Changes in the Unemployment Rate (U.K.)

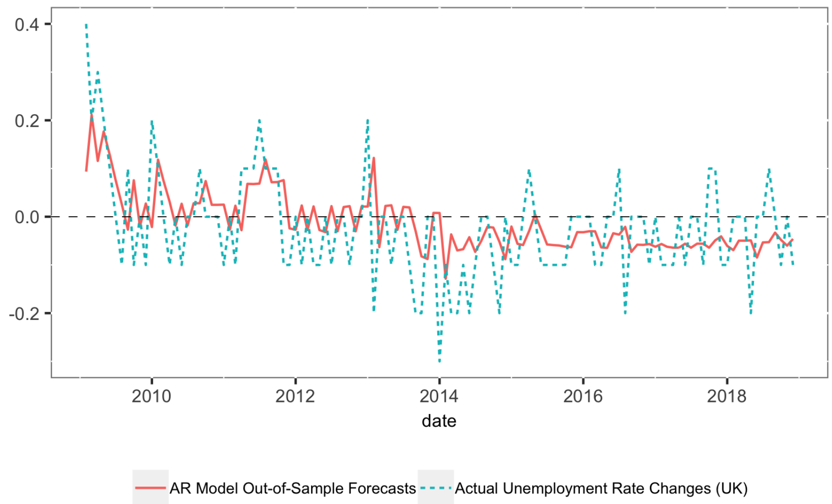


Figure 1.21: Actual and Predicted (Autoregressive Model) Changes in the Unemployment Rate (U.K.)

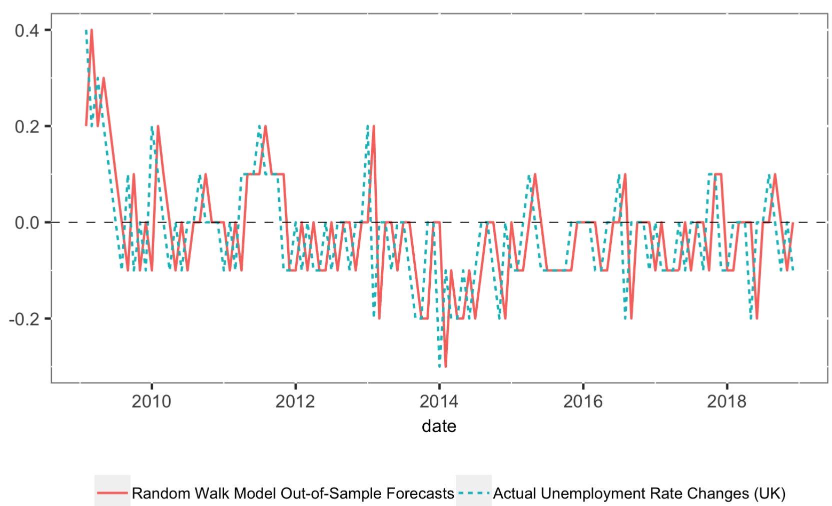


Figure 1.22: Actual and Predicted (Random Walk Model) Changes in the Unemployment Rate (U.K.)

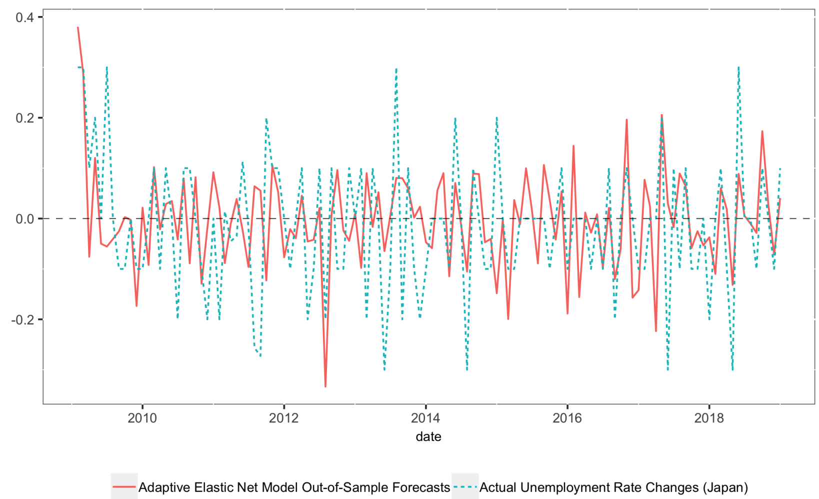


Figure 1.23: Actual and Predicted (Adaptive Elastic-net Model) Changes in the Unemployment Rate (Japan)

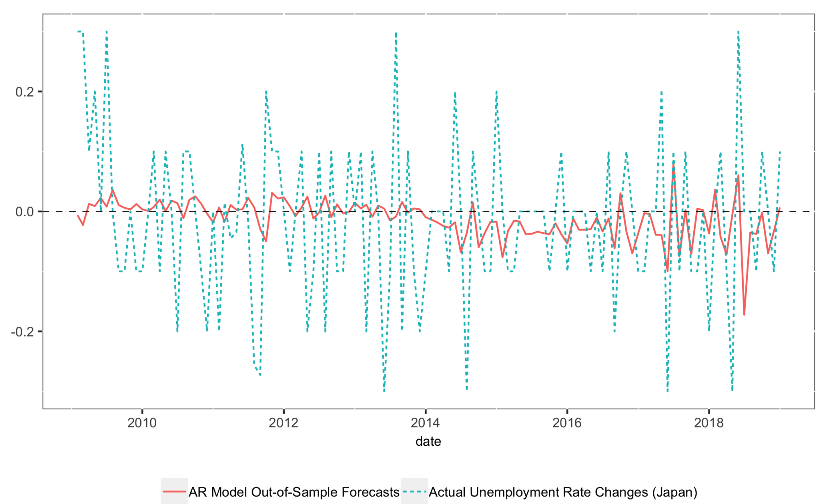


Figure 1.24: Actual and Predicted (Autoregressive Model) Changes in the Unemployment Rate (Japan)

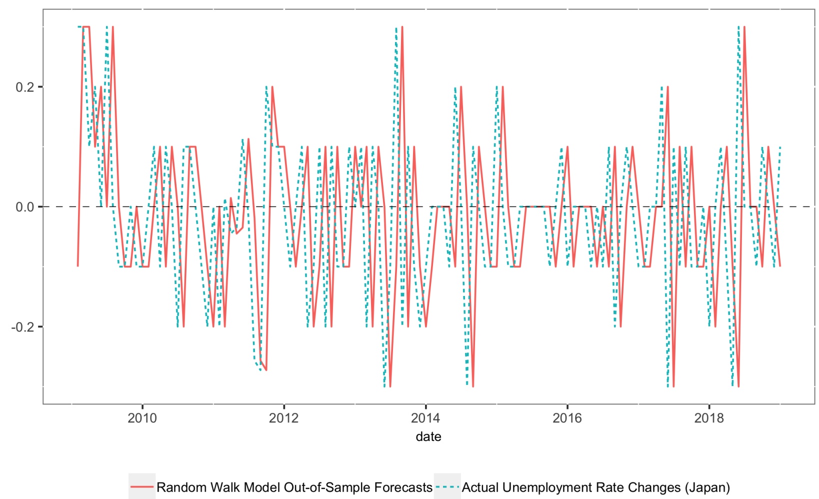


Figure 1.25: Actual and Predicted (Random Walk Model) Changes in the Unemployment Rate (Japan)

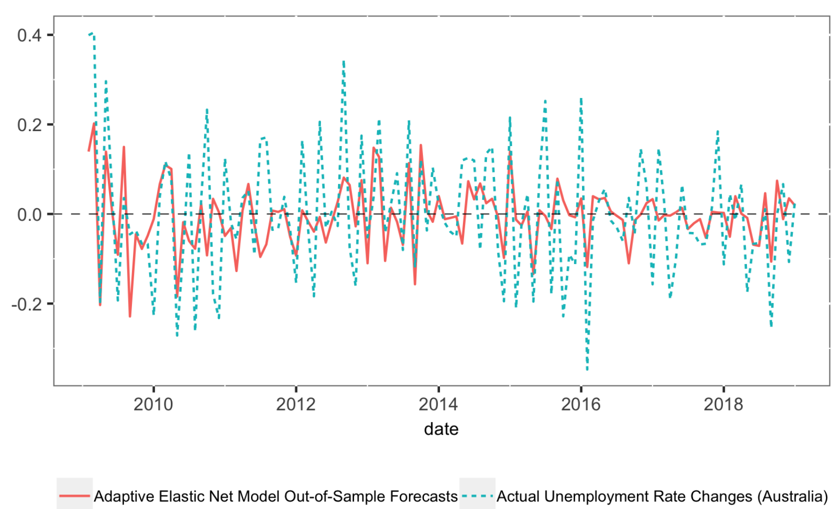


Figure 1.26: Actual and Predicted (Adaptive Elastic-net Model) Changes in the Unemployment Rate (Australia)

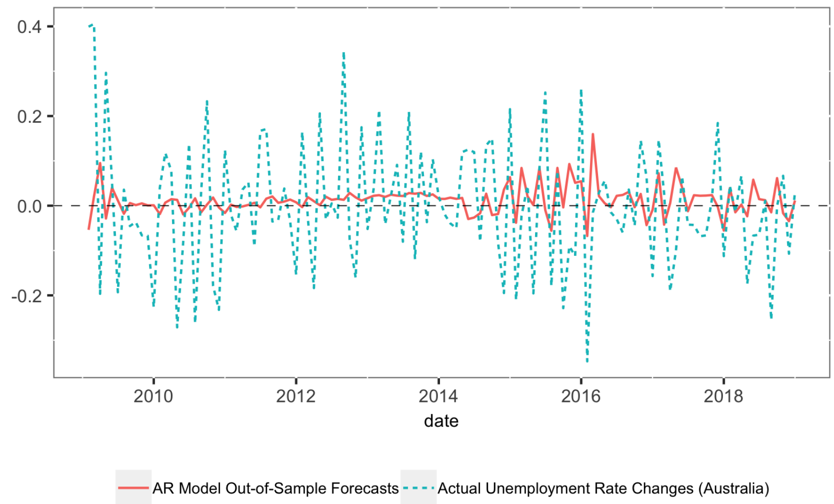


Figure 1.27: Actual and Predicted (Autoregressive Model) Changes in the Unemployment Rate (Australia)

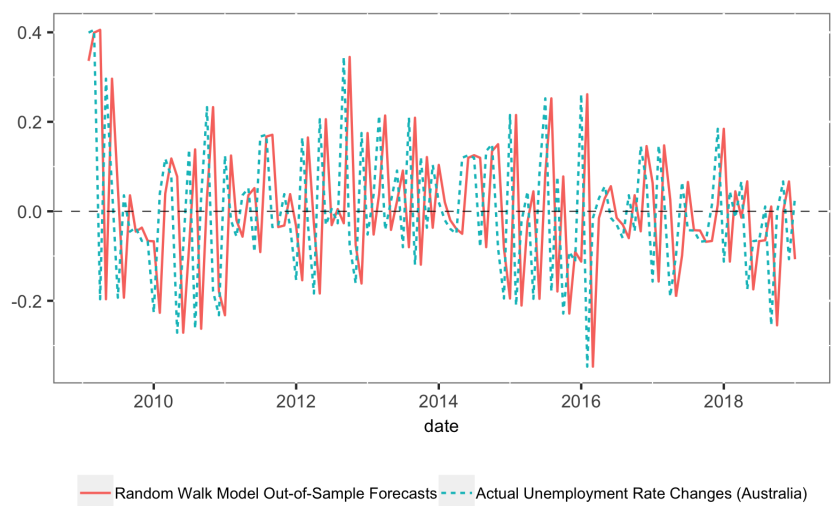
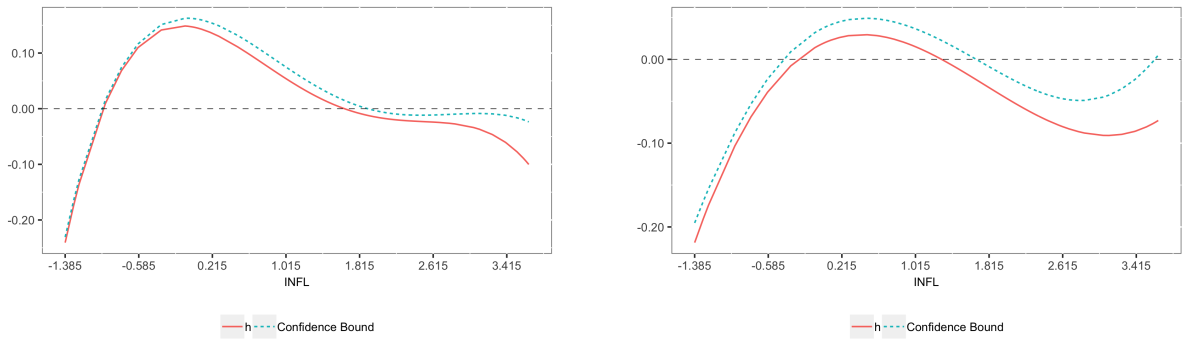


Figure 1.28: Actual and Predicted (Random Walk Model) Changes in the Unemployment Rate (Australia)

Figure 1.29: Forecasting Inflation (U.S.): CSPA

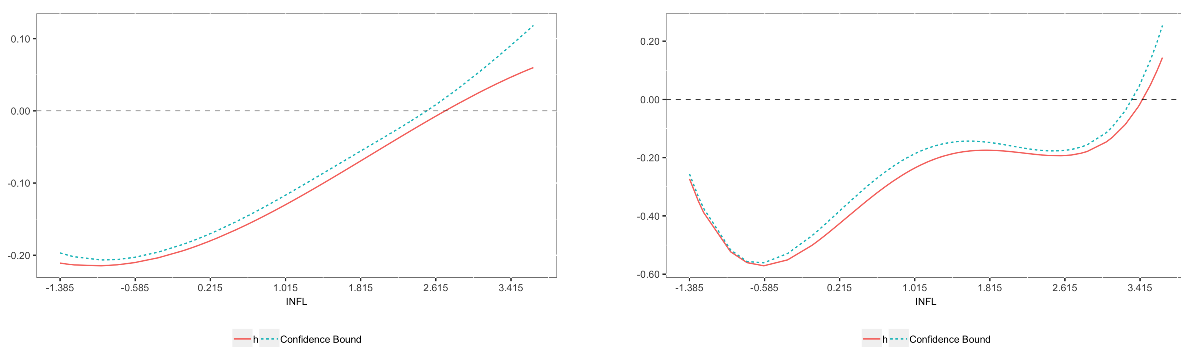


(a) Benchmark: AR

(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in U.S. inflation against the conditioning variable, $x = \text{Inflation}_{US}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.30: Forecasting the Unemployment Rate (U.S.): CSPA

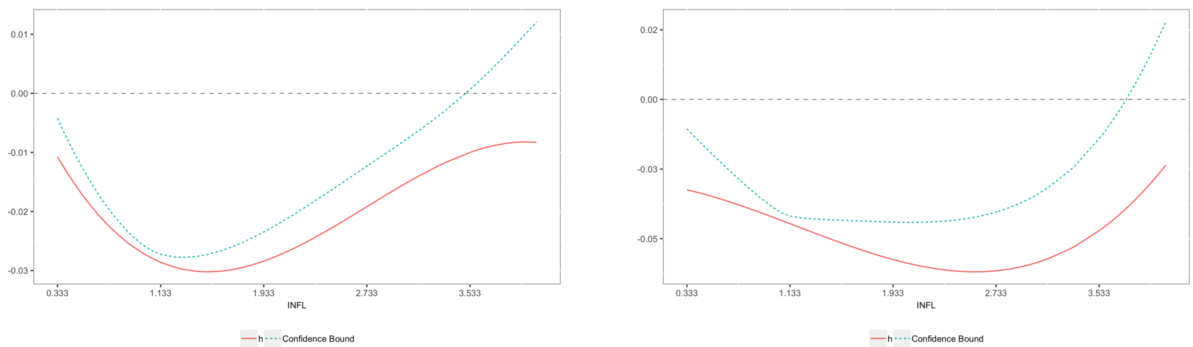


(a) Benchmark: AR

(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in the U.S. unemployment rate against the conditioning variable, $x = \text{Inflation}_{US}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.31: Forecasting Inflation (U.K.): CSPA

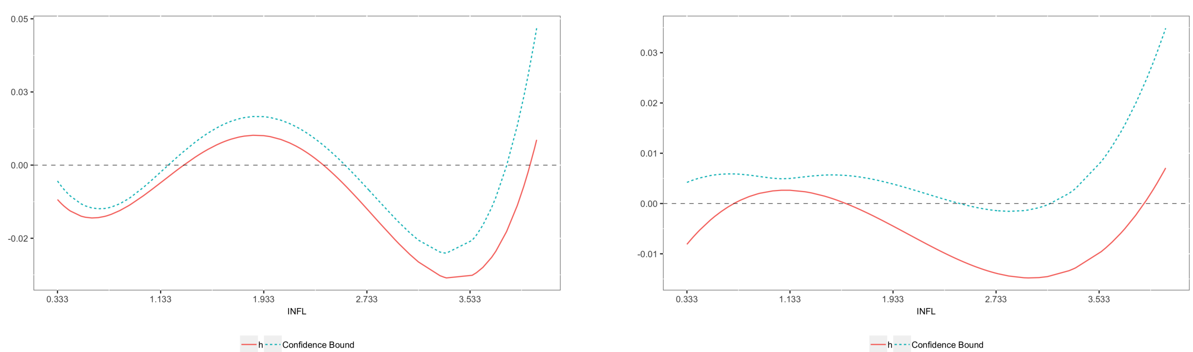


(a) Benchmark: AR

(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in U.K. inflation against the conditioning variable, $x = \text{Inflation}_{UK}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.32: Forecasting the Unemployment Rate (UK): CSPA

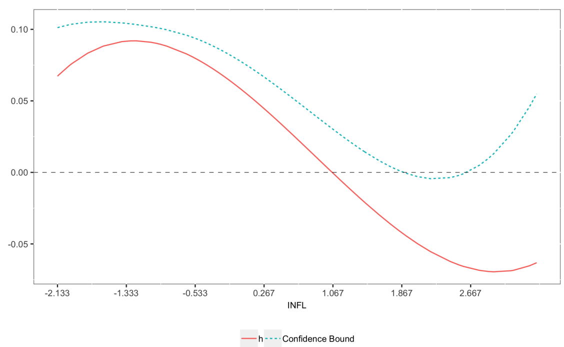


(a) Benchmark: AR

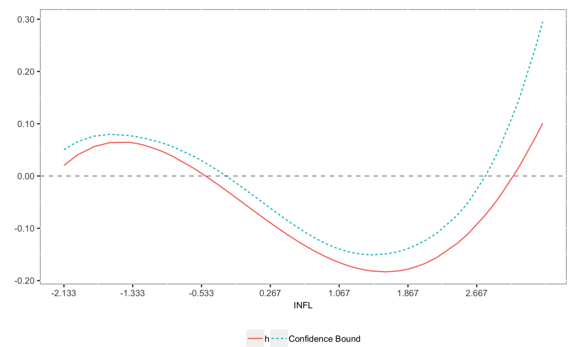
(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in the U.K. unemployment rate against the conditioning variable, $x = \text{Inflation}_{UK}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.33: Forecasting Inflation (Japan): CSPA



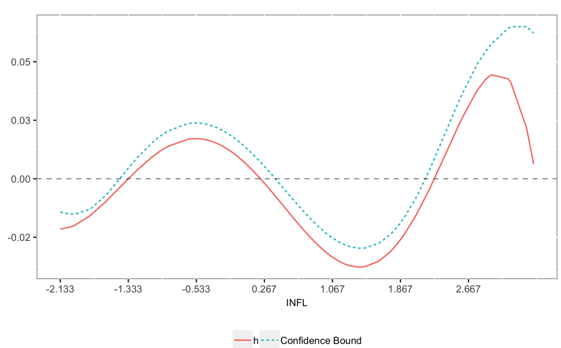
(a) Benchmark: AR



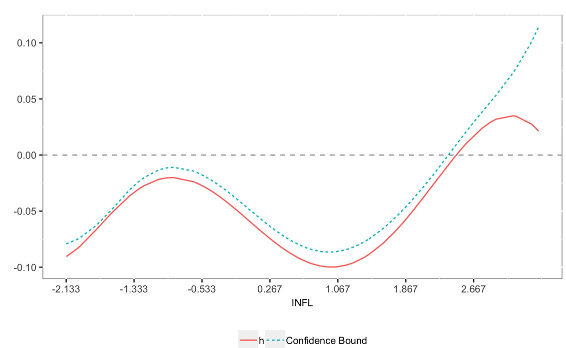
(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in Japan inflation against the conditioning variable, $x = \text{Inflation}_{Japan}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.34: Forecasting the Unemployment Rate (Japan): CSPA



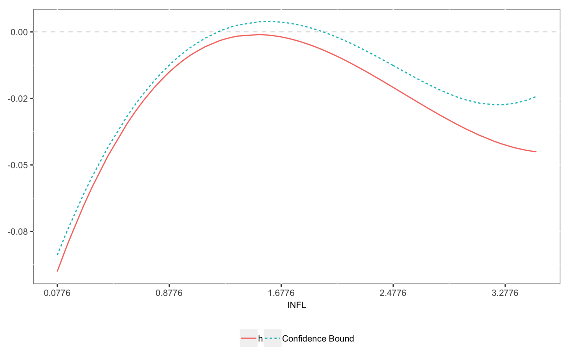
(a) Benchmark: AR



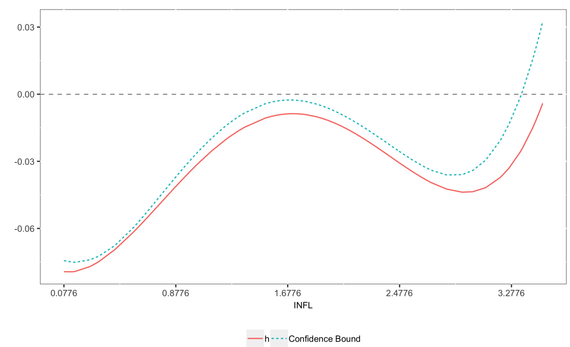
(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in the Japan unemployment rate against the conditioning variable, $x = \text{Inflation}_{Japan}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.35: Forecasting Inflation (Australia): CSPA



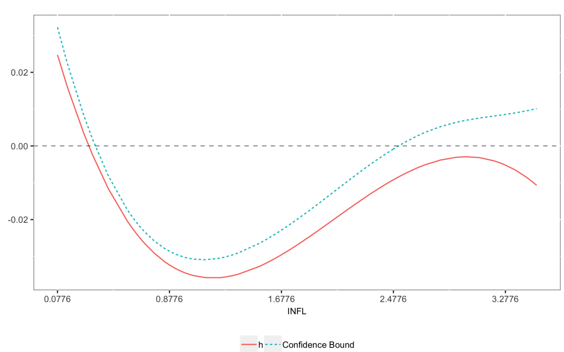
(a) Benchmark: AR



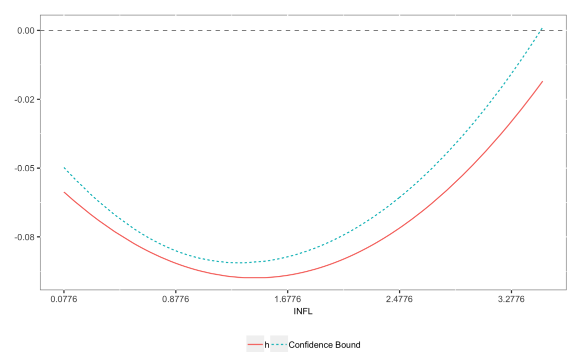
(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in Australia inflation against the conditioning variable, $x = \text{Inflation}_{Australia}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

Figure 1.36: Forecasting the Unemployment Rate (Australia): CSPA



(a) Benchmark: AR



(b) Benchmark: Random Walk

Note: This figure plots the expected mean squared error (MSE) differential between the model with inclusion of Google Trends data and the benchmark models for changes in the Australia unemployment rate against the conditioning variable, $x = \text{Inflation}_{Australia}$. Each panel plots the expected MSE differential against different benchmark models, as well as its 99% upper confidence bound.

1.8 Appendix

1.8.1 Literature on Using Google Search Data in Forecasting/Nowcasting Macroeconomic and Financial Variables

Table 1.6: Use of Google Search Data in Forecasting/Nowcasting Macroeconomic and Financial Variables

Author	Macroeconomic Variable	Country
Giannone, Reichlin, and Small (2008)	GDP	United States
Tatjana, Guenette, and Vasishtha (2017)	GDP	Emerging Markets
Gotz and Knetsch (2019)	GDP	Germany
Askitas and Zimmermann (2009)	Unemployment Rate	Germany
Ross (2013)	Unemployment Rate	United Kingdom
Choi and Varian (2009, 2011)	Unemployment Rate	United States
Scott and Varian (2013)	Unemployment Rate	United States
Reis, Ferreiram, and Perduca (2014)	Unemployment Rate	France
Smith (2016)	Unemployment Rate	United Kingdom
D'Amuri and Marcucci (2017)	Unemployment Rate	United States
Li et al. (2015)	CPI	China
Koop and Onorante (2013)	Macroeconomic Variables	United States
Bok et al. (2018)	Macroeconomic Variables	United States
Da, Engelberg, and Gao (2011)	Stock Prices/Returns	United States
Preis, Moat, and Stanley (2013)	Stock Prices/Returns	United States
Vozlyublennaia (2014)	Stock Prices/Returns	United States
Vlastakis and Markellos (2012)	Stock Market Volatility	United States
hamid and Heiden (2015)	Stock Market Volatility	United States
Goddard, Kita, and Wang (2015)	Exchange Rates	Advanced Economies
Bulut (2015)	Exchange Rates	Advanced Economies
Hasselgren et al. (2018)	Exchange Rates	Emerging Markets

1.8.2 Robustness Check: CW Test with Different Forcasting Horizons

Table 1.7: Point Nowcast Test Results (CW Test)

	Panel A: CW Test against AR Model					
	Δ Unemployment			Δ Inflation		
Country	h = 1	h = 3	h = 6	h = 1	h = 3	h = 6
U.S.	2.047e-3	1.410e-2	5.779e-1	1.609e-1	6.401e-1	4.051e-1
U.K.	1.197e-1	7.240e-1	5.442e-1	3.801e-5	4.220e-2	1.713e-2
Japan	5.686e-2	1.809e-1	8.080e-1	9.660e-3	1.440e-1	9.404e-2
Australia	3.654e-3	1.853e-1	2.844e-2	1.523e-5	8.443e-1	5.437e-2
	Panel B: CW Test against Random Walk Model					
U.S.	1.036e-5	2.602e-2	3.606e-2	2.488e-2	7.860e-1	8.210e-1
U.K.	0.617e-1	1.603e-2	2.030e-1	5.042e-4	3.708e-2	3.332e-1
Japan	1.354e-6	8.605e-2	4.930e-2	2.242e-4	7.664e-1	9.477e-1
Australia	1.602e-5	1.004e-3	1.127e-3	8.565e-6	5.502e-2	7.823e-1

Note: The table reports the p-value of the CW test statistic for each macroeconomic variable. The test uses the Newey-West LRV estimator, which controls for auto-correlation. ***, ** and * represent the 1%, 5% and 10% significance level, respectively.

1.8.3 Construction of the CSPA Test

Let $P(x) = (p_1(x), ..., p_{m_n})^T$ be a vector of approximating basis functions[13]. In the specific empirical test, I choose polynomial series. Coefficients are obtained in the least squares regression:

$$\hat{\boldsymbol{\beta}} \equiv \hat{\mathbf{Q}}^{-1}(\frac{1}{n}\sum_{t=1}^{T} \mathbf{P}(\mathbf{X_t})Z_t), \tag{1.20}$$

and $\hat{\mathbf{Q}} \equiv \frac{1}{n}\sum_{t=1}^{T} \mathbf{P}(\mathbf{X_t})\mathbf{P}(\mathbf{X_t})^T$. Then the functional estimator for $E(Z|\mathbf{X})$ is $h(\cdot) = \mathbf{p}(\cdot)^T\boldsymbol{\beta}$.

Hence $\mu_t = Z_t - h(\mathbf{X_t})$ is the regression error. Define the long-run variance covariance

[13] $P(x)$ contains the constant function ($p_1(x) = 1$)

matrix $(m \times m)$ for $\sqrt{\frac{1}{n}}\sum_{t=1}^{T}\mu_{\mathbf{t}} \otimes \mathbf{P}(\mathbf{X_t})$:

$$\mathbf{A} \equiv Var(\sqrt{\frac{1}{n}}\sum_{t=1}^{T}\mu_{\mathbf{t}} \otimes \mathbf{P}(\mathbf{X_t})).$$

Because the out-of-sample forecasts contain auto-correlation and heteroskedasticity of unknown form and for inference in such models, it is necessary to consider a heteroskedasticity and auto-correlation consistent (HAC) estimator that can consistently estimate the covariance of the model parameters. With the HAC estimator $\hat{\mathbf{A}}$, the variance covariance matrix of the coefficients can be estimated with

$$\hat{\mathbf{\Omega}} \equiv \hat{\mathbf{Q}}^{-1}\hat{\mathbf{A}}\hat{\mathbf{Q}}^{-1}.$$

Therefore, the standard deviation of $\sqrt{\frac{1}{n}}(\hat{h}(x) - h(x))$ can be inferred as

$$\hat{\sigma}(x) \equiv \sqrt{\mathbf{P}(x)^T\hat{\mathbf{\Omega}}\mathbf{P}(x)}.$$

I can use the following implementation procedure to conduct the test:

1. First, I simulate a random vector $\xi \sim \mathcal{N}(0, \hat{\mathbf{\Omega}})$ and let $\hat{t}(x) = \frac{\mathbf{P}(x)^T\xi}{\hat{\sigma}(x)}$.

2. Then simulate a large sample of ξ in the step 1. Set $\bar{K}_n$ to be the $\bar{k}-$ quantile of $\sup_{x\in\mathcal{F}}\hat{t}(x)$ in the sample, where $\bar{k} \equiv 1 - \frac{0.1}{\log T}$. I also set

$$\hat{\Theta} \equiv \{x : \hat{h}(x) \leq \inf_{x\in\mathcal{F}}[\hat{h}(x) + \sqrt{\frac{1}{n}}\bar{K}_n\hat{\sigma}(x)] + 2\sqrt{\frac{1}{n}}\bar{K}_n\hat{\sigma}(x)\}.$$

3. Third, let $\bar{k}_{1-\alpha}-$ be the $(1-\alpha)$ quantile of $\sup_{x\in\hat{\Theta}}\hat{t}(x)$, and the $(1-\alpha)$ upper confidence bound $\hat{\eta}_{1-\alpha}$ is

$$\hat{\eta}_{1-\alpha} = \inf_{x\in\mathcal{F}}[\hat{h}(x) + \sqrt{\frac{1}{n}}\hat{k}_{1-\alpha}\hat{\sigma}(x)].$$

   I reject the null hypothesis with $(1-\alpha)$ significance upper bound if $\hat{\eta}_{1-\alpha} < 0$.

CHAPTER 2

Exchange Rate Forecasting with Taylor Rule Fundamentals and Big Data

2.1 Introduction

Rossi (2013) surveys a vast literature on exchange rate forecasting and suggests that Taylor rule fundamentals have predictive power to forecast nominal exchange rates for the U.S dollar against other advanced economies. Also, real-time data offer better forecasting ability than revised data. However, official data are released to the market participants following a time lag. Therefore, it is worth noting that whether we use real-time or revised data, the market has access to these monthly macroeconomic data in the middle of the following month or later, and the lag in availability will be a problem in the forecasting task.

In this chapter, I use Google Trends search query data that have shown significant nowcasting power for Taylor rule fundamentals in Chapter 1 to forecast monthly nominal exchange rate movements. By using Google Trends data, I am able to obtain a timely description of the state of the economy before official data are released to the public. However, one challenge of using big data such as Google Trends is handling large data sets when the number of potential predictors is larger than the number of observations. In this paper, I investigate the performance of different machine learning models, such as variable selections models, factor models and decision regression trees, in obtaining accurate forecasts of three currency pairs (U.S./U.K., Japan/U.S., and U.S./Australia). In specific, for each exchange rate, I have more than 1,500 potential predictors.

I consider three types of forecasts—point forecasts, unconditional weighted directional forecasts and conditional weighted directional forecasts—evaluate the forecasting performance of the machine learning models over the 2004M1–2019M1 period. By using a 60-month rolling window scheme, I evaluate three types of out-of-sample one-month-ahead forecasts against the random walk without drift. Distinct from exchange rate forecasting with adhoc macroeconomic fundamental data, this chapter aims to forecast exchange rates based on Google search variables that serve as timely proxies for Taylor rule fundamentals.

To evaluate the point forecasts, I implement the Diebold Mariano and West (DMW) test of equal MSEs of the point forecasts and those of the random walk without drift, as well as the Clark West (CW) test of equal predictive ability of our model and the random walk. Due to the presence of auto-correlation in the out-of-sample forecasts, I consider the Newey West long-run variance (LRV) estimator. Across three currency pairs, in most cases the one-month-ahead out-of-sample forecasts of machine learning models beat the random walk model with MSEs that are significantly smaller than those of the benchmark, according to the CW test. And across all the machine learning models, the random forest has the smallest MSE of other models and performs significantly better than the random walk.

The weighted directional forecasts that are relevant to the profitability of my forecasts are better than typical binomial tests that ignore the magnitude of changes in subsequent exchange rates. According to the weighted directional test proposed by Kim, Liao, and Tornell (2014), the test statistic shows significantly better performance for all currencies at the 10% significance level than the benchmark. However, the performance varies across different empirical models. The adaptive elastic net model is proven to outperform the random forest with the highest success ratio of its directional forecasts. Also, the random forest shows the highest t-statistic of weighted directional forecasts for the U.S./Australia pair. Based on the varying performance of the models, I ask further when we can rely on Google search data for exchange rates.

Therefore, I consider the performance of these weighted directional forecasts conditional on some relative cyclical macroeconomic variable that can describe the relative state of the economies. Unlike the previous unconditional weighted directional forecasts, in terems of their unconditional moment performance against the random walk, the conditional evaluation enables us to know when exactly my out-of-sample forecasts are better than betting on the random walk with zero profit. Twisting the conditional superior predictive ability test by Li, Liao, and Quaedvlieg (2019), the null hypothesis says that the conditional expected weighted directional forecasts are lower than zero across all conditioning states of a cyclical variable. Therefore, under the alternative, there exist certain states in which my out-of-sample weighted directional forecasts are profitable. I choose the inflation differential of two countries as the conditioning variable. In the Appendix, I also consider the GDP growth differential for the robustness check. All of the currency one-month-ahead weighted directional forecasts reject the null hypothesis of zero profit. Specifically, I see that the weighted directional forecasts are significantly positive on the tails of the conditioning variable distribution.

The rest of the paper is organized as follows. Section 2.2 reviews the related literature review on exchange rate forecasting. In Section 2.3, I replicate the findings of Molodtsova and Papell (2009). Section 2.4, I construct the out-of-sample forecasts for three major currencies using a variety of machine learning models. Section 2.5 presents the point forecasts. Section 2.6 gives the results of both unconditional and conditional weighted directional forecasts, and Section 2.7 concludes.

2.2 Literature Review

First, I use Google Trends data for exchange rate forecasting that ties exchange rate movements to Taylor rule fundamental variables. In specific, I aim to capture people's collective perceptions of macroeconomic variables, such as inflation and the unemployment rate, by analyzing search engine query volume data from Google Trends. Since the seminal work of Meese and Rogoff (1983), it has been argued that the random walk generates better exchange rate out-of-sample forecasts than any other economic model. For the U.S. and other advanced economies, central banks are advised to set a short-term nominal interest rate as a function of the inflation rate, the output gap, and other constant variables. These variables have been described as "Taylor rule fundamentals." Recent literature has indicated that

Taylor-rule fundamentals have predictive power to forecast nominal exchange rates for the U.S. dollar against other advanced economies. Molodtsova and Papell (2009) and a recent survey by Rossi (2013) find evidence that these variables can be used to produce forecasts of the exchange rate that outperform the random walk forecast (i.e., of no change in the log of the exchange rate) at short horizons of one month by using the Clark and West (2006) test to compare model predictions against the random walk.

Although several studies including those by Engel et al. (2008), Molodtsova et al. (2008), Molodtsova et al. (2011), Wang and Wu (2012), and Binici and Cheung (2012) advocate using Taylor-rule fundamentals to predict exchange rates, their results rely on the availability of timely macroeconomic variables such as inflation and GDP. However, official government data release with a time lag. However, in the era of the Internet, big data have emerged as an important addition to traditional data sources. As shown in Chapter 1, Internet search data are able to provide more accurate nowcasts for economic indicators than other conventional methods. Bulut (2015) shows that Google Trends-based forecasts are better for predicting directional changes in monthly nominal exchange rates after the 2008 financial crisis. Markiewicz1 et al. (2018) measure investor attention to different sources of economic information using search intensity as measured by Google Trends, and apply machine learning models to show predictability in forecasting exchange rates movements. Also, Chojnowski and Dybka (2017) extend the present value model based on observable fundamentals by including three unobserved fundamentals sentiments extracted from Google Trends for different markets.

Second, in this chapter, I consider a variety of machine learning models to either select or form important predictors in the environment of rich data. Specifically, I consider a number of variable selection models in choosing relevant search queries from thousands of candidate predictors. Dunis and Williams (2002), Brandl et al. (2009), and Plakandaras (2015) see an improved out-of-sample forecasting ability of dimension-reduction models compared with alternative forecasting models. Then I implement a dynamic factor model to form dynamic parsimonious and suitable representation of Google Trends. Rather than constructing static common factors using such methods as principle component analysis (Engel, Mark and West, 2014), I recognize the serial correlation among the Google Trends. Bok et al. (2017) and Chernis (2017) find that the dynamic factor model outperforms univariate benchmarks in forecasting short-term macroeconomic variables. Moreover, I consider using some "black box" models such as decision tree regression and random forest model to conduct the out-of-sample forecasting. For out-of-sample performance, machine learning algorithms such as random forest can do significantly better than ordinary least squares, even at moderate sample sizes and with a limited number of covariates. Pradeepkuma and Ravi (2016) use a quantile regression random forest model and find a smaller mean absolute percentage error for the out-of-sample major exchange rate forecasts than other machine learning models.

Third, regarding out-of-sample forecast accuracy ability, there have been mixed findings in the literature. Since Meese and Rogoff (1983), it has been difficult for fundamentals-augmented point forecasts to beat the random walk, with few exceptions. The null hypothesis of the CW test has been rejected by several authors such as Gourinchas and Rey (2007), using net foreign assets, and Molodtsova and Papell (2009), using Taylor rule fundamentals. However, few papers focuse on directional forecasts. I conduct a directional forecast test proposed by Kim et al. (2019) that weights each directional forecast by the subsequent exchange rate change, and so gives more weight to directional forecasts associated with bigger exchange rate moves. Moverover, this paper conciders not only unconditional superior predictive accuracy, but explores *when* weighted directional forecasts outperform the random walk. Complementing the theoretical framework of the conditional superior predictive of Li, Liao, and Quaedvlieg (2019), the test enables us to know when to use forecasts from which machine learning model conditional on a relative economic indicator.

2.3 Exchange Rate Forecasting with Taylor Rule Fundamentals

Suppose the Fed and other central banks follow the Taylor rule to set the short-term nominal interest rate

$$i_t = \mu + \lambda \pi_t + \gamma y_t \tag{2.1}$$

$$i_t^* = \mu^* + \lambda^* \pi_t^* + \gamma^* y_t^* \tag{2.2}$$

where π_t is the inflation rate, y_t is the output gap, and (μ, λ, γ) are constants. Sometimes we can replace y_t with the unemployment gap. If uncovered interest rate parity holds, the optimal forecast for the change in the exchange rate between t and $t+1$ is equal to the short-term interest rate between two countries. Then, under uncovered interest rate parity between the U.S. and the foreign country,

$$E_t(s_{t+1} - s_t) = i_t - i_t^* \tag{2.3}$$

where s_t is the log of the dollar price of the foreign currency. If UIP holds, by substituting equation (2.1) for equation (2.3), I would obtain

$$E_t(s_{t+1} - s_t) = \mu + \lambda \pi_t + \gamma y_t - (\mu^* + \lambda^* \pi_t^* + \gamma^* y_t^*). \tag{2.4}$$

Intuitively, an increase in the inflation rate or the unemployment gap of the U.S. will increase the U.S. interest rate and predict an immediate appreciation in the dollar. However, official macroeconomic variable data releases follow a lag: Most macroeconomic data are released in the middle or latter half of the next month. Although Molodtsova and Papell (2009) and others find promising out-of-sample predictive power with Taylor rule fundamentals, their results rely on exante official macroeconomic data that are not available at the time of making predictions in exchange rate movements.

In Section 2.3, I replicate the findings of Molodtsova and Papell (2009) with three major exchange rates: the USD vis-à-vis the pound sterling , the Japanese yen, and the Australian dollar with the sample period from 2004M1 to 2019M1. Price level data are the consumer price index from the IMF's *International Financial Statistics* (IFS). Unemployment rates are from the OECD Original Release and Revisions Database. To construct one-month-ahead forecasts of changes in the exchange rate between time t and $t+1$, I estimate the following regression equation[1] with rolling regressions with a 60-month window:

$$\Delta s_{t+1} = b_0 + b_1\pi_t + b_2\pi_t^* + b_3 y_t + b_4 y_t^* + b_5 i_{t-1} + b_6 i_{t-1}^* + \epsilon_t \tag{2.5}$$

Then the one-month-ahead out-of-sample forecast is produced using the above estimated equation. The first forecast starts from 2009M3 and the last ends in 2019M1. Based on all out-of-sample forecasts[2], I am able to compute the out-of-sample MSE of the Taylor rule model and compare to the random walk. I report the MSE ratios of the Taylor rule model and the random walk model in Table 2.1. In the three currency pairs, all MSE ratios are lower than one, which means that the Taylor rule model produces better forecasts than the random walk benchmark. To evaluate the statistical significance, I report the CW test statistic and the associated p-values in Table 2.1. It shows that the Taylor rule model significantly outperform the random walk model at a 1% significance level for the three exchange rates. Note that the CW test corrects for the fact that the Taylor rule model is nested with the random walk model, which does not require parameter estimation. In the Appendix, I also report the DMW test statistic and the corresponding p-values in Table 2.7. I can see that it becomes harder to reject the null hypothesis with the DMW test, which has a smaller test statistic. However, Taylor rule models are still significantly better than the random walk model at 1% or 5% significance levels.

[1] I augment Taylor rule fundamental variables with lagged interest rates that are shown to be the most successful by Molodtsova and Papell (2009) and Engel et al. (2018).

[2] There are 118 out-of-sample forecasts for each currency pair.

Table 2.1: Taylor Rule Model: Replication of Molodtsova and Papell (2009)

Currency	MSE (Taylor rule model)	MSE (random walk)	MSE ratio	CW test statistic (p-value)
U.S./U.K.	$1.43e^{-3}$	$7.8e^{-4}$	0.547	$3.444***$ ($2.86e^{-4}$)
Japan/U.S.	$1.39e^{-3}$	$9.17e^{-4}$	0.656	$3.408***$ ($2.86e^{-4}$)
U.S./Australia	$1.53e^{-3}$	$8.00e^{-4}$	0.522	$3.513***$ ($2.22e^{-4}$)

Note: This table shows MSE for both Taylor rule forecasts and random walk model forecasts for three currency pairs. It also reports CW test statistic and its respective $p-$value. $*$, $**$, $***$ indicate that the alternative model significantly outperforms the random walk at 10%, 5%, and 1% significance level, respectively, based on standard normal critical values for the one-sided test. Forecast results are based on out-of-sample forecasts starting from 2009M3 to 2019M1 for all currency pairs.

2.4 Machine Learning Methods with High-dimensional Covariates

Note that the macroeconomic data at time t are not announced until the middle or the latter half of the month $(t+1)$. Therefore, these significant results are based on the exante macroeconomic data. I introduce Taylor rule fundamentals-related search query data from Google Trends to forecast exchange rate movements. In Section 2.4, I consider a variety of machine learning methods to handle the high-dimensional covariates space. Specifically, I consider variable selection models, the dynamic factor model, and random forest and Bagging methods.

In Chapter 1, I see the potential value of using Google Trends search to proxy the current values of Taylor rule fundamentals variables in four countries. I can also exploit them in exchange rate forecasting. In this chapter, instead of inserting a substitute for the current state of official government data in the specification (2.1), I aim to directly explore these Taylor rule fundamentals-related search data in exchange rate forecasting.

In this section, I consider a variety of machine learning models to forecast exchange rate movements. If we define $\mathcal{G}$ as the matrix of the Taylor rule related Google Trends data for the U.S. and $\mathcal{G}^*$ for the foreign country, I consider the following forecasting model:

$$s_{t+1} - s_t = \mathbf{F}(\mathcal{G}_t, \mathcal{G}_t^*) \tag{2.6}$$

where s_t is the end-of-month nomical exchange rate[3] and $\mathbf{F}(\cdot)$ is the functional form for the machine learning models I consider below.

2.4.1 Variable Selection Model

In Chapter 1, I introduced the class of variable selection or shrinkage models to nowcast macroeconomic fundamental variables. In particular, the function form in Equation 2.6 is a linear model with a regularization (penalty) function. Chapter 1 presents a detailed explanation of the models. Here, I consider a ridge regression (with a quadratic l_2 penalty), adaptive LASSO (with a weighted l_1 penalty), and adaptive elastic-net model (with a weighted combination of l_1 and l_2 penalty). Also, I consider the LARS algorithm (Efron et al., 2004) to obtain the entire solution, and parameters in front of the penalty terms are chosen with five fold cross-validation.

2.4.2 Dynamic Factor Model

The dynamic factor model has received considerable attention for its prediction ability for a data set when the number of time-series covariates exceeds the number of observations. Introduced by Geweke (1977), a growing body of empirical research (Giannone et al., 2004; Watson, 2004; and Bai and Ng, 2008) shows that a few factors can explain the majority of variance in macroeconomic variables that include GDP, employment, and prices. In the exchange rate forecasting literature, Engel et al. (2012) construct factors from exchange rates and use deviations from the factors to forecast the exchange rates themselves.

In this subsection, I consider constructing factors from the high-dimensional Google Trends search query data set and use these factors to determine exchange rate movements.

[3] I obtain the data from the Federal Reserve Bank of St. Louis (FRED) database.

Following Stock and Watson (2010), I assume m latent factors in Google Trends search data following a time-series process, and the dynamic factor model is

$$\mathcal{G}_t = \lambda(L)M_t + \epsilon_t \tag{2.7}$$
$$M_t = \phi(L)M_{t-1} + \nu_t$$

where there are M Google Trends, so $\mathcal{G}_t$ and ϵ_t are $M \times 1$. Assume that there are m dynamic factors so that f_t and ν_t are $m \times 1$,and L is the lag operator. Then the forecasting function **F** in 2.6 follows the specifications in system 2.7. Inference of the dynamic factor models can be conducted using Kalman filter techniques to handle the nonlinearity of the data. Since I expand the Google Trends query data based on seeding words, I expect high collinearity in the covariates space. Then, in practice, I first cluster the Google Trends data with a simple k-means algorithm before feeding it into the dynamic factor model. This step is reasonable and provides efficiency in the inference.

2.4.3 Decision Trees and Random Forests

A decision tree, also called a regression tree, is a prediction model based on recursively partitioning the covariate space. The decision tree is built using a randomly selected data sample or randomly selected covariates from the original data to split each node which leads to good out-of-sample predictions.

To demonstrate the use of regression trees, I graph a tree that predicts U.S./U.K. exchange rate movement using more than 1,000 Taylor rule fundamentals-related Google Trends search query data. The tree is shown in Figure 2.1 and the rules are indicated above the nodes. This example provides a simple demonstration of important predictors for some of the exchange rate movement values. In this example, the root regressor starts with query time series[4] "high food prices." If "guarantee pay" < 0.625, "number of people unemployed" ≥ 0.03, and "rail fare" < -0.1, then the model predicts that the log nominal exchange rate

[4]The covariates space is preprocessed by extracting out seasonality, first-differenced, and normalized with mean 0 and standard deviation 1.

of the U.S/U.K. goes down by -0.0133. I can also interpret that among other Google Trends queries, "high food prices," "guarantee pay," "number of people unemployed," and "rail fare" are important predictors of that particular value of exchange rate movement. One problem

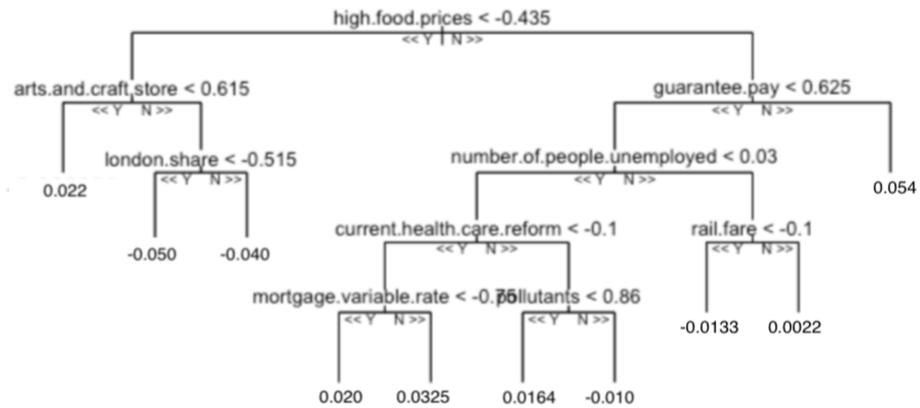


Figure 2.1: A Simple Decision Tree for U.S/U.K. Exchange Rate Movements with Google Trends Covariates

with a single regression tree is that they tend to "overfit" the data and lead to high variances in out-of-sample forecasts. However, one can create more robust predictions by bootstrapping multiple subsamples, fitting a regression tree to each subsample, and then averaging the predictions across the bootstrapped samples. A random forest is then a collection of regression trees and each of which is in a bootstrapped sub-sample of the data set.

2.5 Point Forecasts Performance

Our data cover 2004M1–2019M1. I use the same set of Google Trends search queries as used in Chapter 1 for macroeconomic nowcasting. The covariates consist of about 1,500 Google Trends search queries that are relevant to the macroeconomic variables, such as unemployment rate and inflation for home and foreign countries for each exchange rate.

All covariates are seasonally adjusted, first-differenced and normalized with mean zero and standard deviation one. I use the first 60 observations for in-sample estimation and obtain out-of-sample forecasts from 2009M3 to 2019M1 in a rolling-window scheme. In this section, I describe the main results of my analysis. I report point forecast performance of five machine learning models based on the out-of-sample MSE compared with that of the random walk model. Also, Tables 2.2-2.4 report CW and DMW test statistics with their associated p-values to show whether the Google Trends forecasts outperform the random walk. Also, each table ranks machine learning models by their MSE ratio[5].

The results indicate that the majority of the machine learning models produce forecasts that have lower MSEs than that of the random walk benchmark. In Table 2.2, all of the models, except for the ridge model, have MSE ratios of less than one. According to the CW test, their out-of-sample forecasts are significantly better than that of the benchmark. It is also worth noting that the dynamic factor model and random forest even reject the null hypothesis with the DMW test at the 5% and 10% significant level, respectively. For the Japan/U.S. exchange rate (Table 2.3), all of the Google Trends forecasts have smaller MSEs than the random walk, and the random forest has the smallest MSE ratio, 0.609. Similarly, the random forest dominates the other models in forecasting the U.S./Australia pair in Table 2.4, and the null hypothesis is rejected with the DMW test. The results also imply that random forest—which is a solution to reduce the variance of decision trees by bootstrapping aggregation of randomly constructed regression trees—shows consistently dominant performance over other models, such as variable selection and dynamic factor models.

[5]MSE ratio is the ratio of the MSE of the machine learning model over the random walk.

Table 2.2: Point Forecasting Performance for the U.S./U.K.

Model	MSE	MSE ratio	CW test statistic (p-value)	DMW test statistic (p-value)
Dynamic Factor Model	$0.844e^{-3}$	0.631	$3.084***$ $(1.01e^{-3})$	$2.275**$ $(1.327e^{-2})$
Random Forest	$1.007e^{-3}$	0.705	$2.637***$ $(4.181e^{-3})$	$1.611*$ $(5.620e^{-2})$
Adaptive Elastic Net	$1.182e^{-3}$	0.828	$2.030**$ $(2.117e^{-2})$	0.821 $(2.074e^{-1})$
Adaptive LASSO	$1.164e^{-3}$	0.839	$2.003**$ $(2.260^{e}-2)$	0.713 $(2.393e^{-1})$
Ridge	$1.703e^{-3}$	1.039	1.061 $(1.44e^{-1})$	-0.123 $(5.489e^{-1})$

Note: This table shows MSE for five machine learning forecasts and random walk model forecasts for the U.S./U.K. exchange rate. It also report the CW test statistic (its respective p-value) and the DMW test statistic (its respective $p-$value). $*$, $**$, $*$ indicate that the alternative model significantly outperforms the random walk at 10%, 5%, and 1% significance level, respectively, based on standard normal critical values for the one-sided test. Forcast results are based on out-of-sample forecasts from 2009M3 to 2019M1, and the table is ranked by MSE value.

Table 2.3: Point Forecasting Performance for the Japan/U.S.

Model	MSE	MSE ratio	CW test statistic (p-value)	DMW test statistic (p-value)
Random Forest	$8.630e^{-4}$	0.609	$3.800***$	$2.250**$
			$(7.230e^{-5})$	$(1.408e^{-2})$
Adaptive Elastic Net	$1.017e^{-3}$	0.727	$2.721***$	1.333
			$(3.258e^{-3})$	$(9.386e^{-2})$
Adaptive LASSO	$1.090e^{-3}$	0.779	$3.589***$	1.234
			$(1.65e^{-4})$	$(1.110e^{-1})$
Ridge	$1.096e^{-3}$	0.783	$2.899***$	1.071
			$(1.869e^{-3})$	$(1.443e^{-1})$
Dynamic Factor Model	$1.418e^{-3}$	0.986	$1.577*$	0.054
			$(5.74e^{-2})$	$(4.784e^{-1})$

Note: **Note:** This table shows MSE for five machine learning forecasts and random walk model forecasts for the Japan/U.S. exchange rate. It also report the CW test statistic (its respective p-value) and the DMW test statistic (its respective $p-$value). $*$, $**$, $*$ indicate that the alternative model significantly outperforms the random walk at 10%, 5%, and 1% significance level, respectively, based on standard normal critical values for the one-sided test. Forcast results are based on out-of-sample forecasts from 2009M3 to 2019M1, and the table is ranked by MSE value.

Table 2.4: Point Forecasting Performance for the U.S./Australia

Model	MSE	MSE ratio	CW test statistic (p-value)	DMW test statistic (p-value)
Random Forest	$7.980e^{-4}$	0.571	$3.493***$	$2.106**$
			$(2.390e^{-4})$	$(1.972e^{-2})$
Adaptive Elastic Net	$1.057e^{-3}$	0.756	$2.721***$	1.333
			$(3.258e^{-3})$	$(9.386e^{-2})$
Adaptive LASSO	$1.090e^{-3}$	0.779	$2.746***$	1.073
			$(3.012e^{-3})$	$(1.438e^{-1})$
Ridge	$1.098e^{-3}$	0.785	$2.746***$	0.897
			$(3.019e^{-3})$	$(0.187e^{-1})$
Dynamic Factor Model	$1.812e^{-3}$	1.296	0.581	-0.884
			$(2.806e^{-1})$	$(8.099e^{-1})$

Note: This table shows MSE for five machine learning forecasts and random walk model forecasts for the U.S./Australia exchange rate. It also report the CW test statistic (its respective p-value) and the DMW test statistic (its respective $p-$value). $*$, $**$, $*$ indicate that the alternative model significantly outperforms the random walk at 10%, 5%, and 1% significance level, respectively, based on standard normal critical values for the one-sided test. Forcast results are based on out-of-sample forecasts from 2009M3 to 2019M1, and the table is ranked by MSE value.

2.6 Weighted Directional Forecasts by Subsequent Exchange Rate Changes

In this subsection, I am interested in testing whether the weighted directional forecasts which are defined as the multiplication of directional forecasts and the magnitude of changes in subsequent exchange rates—outperform random walk forecasts. I employ the weighted directional test proposed by Kim, Liao, and Tornell (2014). With directional forecasts $\{D_t^{ji}\}_{t=n_0^{ji}}^{n^{ji}}$ of each exchange rate i from the machine learning model j, I consider the following relative profitability measure:

$$D_t^{ji}(s_{t+1}^i - s_t^i). \tag{2.8}$$

Therefore, $D_t^{ji}(s_{t+1}^i - s_t^i)_{t=n_0^{ji}}^{n^{ji}}$ are series of weighted directional forecasts for currency i from model j. Note that under the benchmark random walk model, the optimal forecast is a zero exchange rate change, namely, $D_t^{0i}(s_{t+1}^i - s_t^i)_{t=n_0^{0i}}^{n^{0i}} = 0 \quad \forall i, t$.

2.6.1 Unconditional Weighted Directional Forecasts Test

First, I report the forecast success ratios of all five machine learning models for three currency pairs in Table 2.5. In the table, all of the exchange rates have success ratios over 50% for one-month-ahead forecasts. Of the machine learning models, the adaptive elastic-net model shows the highest success ratio.

Table 2.5: Directional Forecast Success Ratios

Model	U.S./U.K.	Japan/U.S.	U.S./Australia
Adaptive Elastic Net	0.602	0.639	0.529
Random Forest	0.576	0.588	0.580
Adaptive LASSO	0.542	0.521	0.597
Ridge	0.585	0.521	0.580
Dynamic Factor Model	0.534	0.546	0.563

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts. Out-of-sample forecasts are from 2009M3 to 2019M1.

Then I formally test the performance of directional forecasts against the random walk model. Specifically, I am interested in the weighted directional forecasts in Equation 2.8. The details of the test can be found in are laid out in Kim, Liao and Tornell (2014). In Table 2.6, the performance varies with models and currencies, even though all out-of-sample directional forecasts have more than 50% success ratios indicated in Table 2.5. The adaptive elastic net forecasts show consistently significantly better performance for all currencies. And random forest forecasts perform signigicantly better for the U.S./U.K. and U.S./Australia. Note that each model of the five machine learning models significantly outperforms the benchmark at least for one currency. And sometimes one model is better than another according to the $t-$statistics. Then it is natural to ask *when* a particular model is more preferrable than another. Especially, when we face a list of candidate forecast models, a rigorous evaluatation method of their relative performance is much desirable.

Table 2.6: Unconditional Weighted Directional Forecasts

Model	U.S./U.K.	Japan/U.S.	U.S./Australia
Adaptive Elastic Net	1.570$*$	1.230$*$	1.520$*$
	(0.058)	(0.100)	(0.065)
Random Forest	1.480$*$	0.607	2.050 $**$
	(0.069)	(0.272)	(0.020)
Adaptive LASSO	1.530$*$	0.844	0.838
	(0.063)	(0.199)	(0.201)
Ridge	2.070 $**$	0.045	0.801
	(0.019)	(0.482)	(0.211)
Dynamic Factor Model	1.010	1.250$*$	0.829
	(0.156)	(0.099)	(0.204)

Note: This table shows the $t-$statistics for the weighted directional forecasts for each currency and each machine learning model. $**$,$**$ and $*$ represent the 1%, 5% and 10% significance level, repectively. And the out-of-sample forecasts start from 2009M3 to 2019M1.

2.6.2 Conditional Weighted Directional Forecasting Superior Test

In this subsection, I consider the conditional superior ability test proposed by Li, Liao, and Quaedvlieg (2018) to assess predictive performance based on a cyclical indicator variable. Since the exchange rate is a relative price, I am curious regarding how the weighted directional forecasts perform conditional on two countries' relative conditions. If

$$D_t^{0i}(s_{t+1}^i - s_t^i)_{t=n_0^{0i}}^{n^{0i}} > 0 \quad \forall t,$$

then the random walk model with zero profit is beaten by the weighted directional forecasts. Therefore, I consider the null hypothesis to be

$$H_0 : E(D_t^{ji}(s_{t+1}^i - s_t^i)|X_t = x_t) \leq 0 \quad \text{almost surely, } t = 1, 2, ..., \tag{2.9}$$

where X_t is the conditional variable that captures the state of one country or the relative levels of both countries. The null hypothsis says that random walk weighted directional forecasts outperform the forecasts of a competing machine learning model uniformly across all states in the information set $\mathcal{F}$. To test the conditional moment inequality, I must estimate conditional expected profit using functional inference. With the intersection-bound method, the null hypothesis can be written as

$$H_0 : \eta \equiv \sup_{x \in \mathcal{F}} E(D_t^{ji}(s_{t+1}^i - s_t^i)|X_t = x_t) \leq 0 \tag{2.10}$$

where $\mathcal{F}$ is the information set. The $(1 - \alpha)$ lower confidence bound can be constructed as follows:

$$\lim_{n \to \infty} \sup \mathrm{P}(\eta \geq \hat{\eta}_{1-\alpha}) \geq 1 - \alpha. \tag{2.11}$$

The alternative test rejects the null when $\hat{\eta}_{1-\alpha} > 0$, with the type I error bounded by α in a sufficient large sample.

Constructing the CSPA test involves estimating the conditional expected profit function nonparametrically with least squares regressions. In this empirical exercise, I consider a conditional variable such as the moving average inflation rate differential between both countries. Then I estimate the expected MSE differential function nonparametrically with polynomial series as the approximating basis functions. The number of polynomials is chosen using minimum Akaike Information Criteria (AIC) using five fold cross-validation. I lay out the details of implementation of the CSPA test in the following. Because of the right tail of the distribution of the conditional variable, the lower confidence bounce becomes wildly. By trimming the right tail of the distribution by 5%, I focus on the truncated range of the conditional variable.

Let $P(x) = (p_1(x), ..., p_{m_n})^T$ be a vector of approximating basis functions[6], and denote $Z_t^{ji} \equiv D_t^{ji}(s_{t+1}^i - s_t^i)$. In the specific empirical test, I choose polynomial series. For each currency i and machine learning model j, I regress Z_t^{ji} onto the polynomial basis series. Coefficients are obtained using least squares regression:

$$\hat{\boldsymbol{\beta}} \equiv \hat{\mathbf{Q}}^{-1}(\frac{1}{n}\sum_{t=1}^{T}\mathbf{P}(\mathbf{X_t})Z_t), \tag{2.12}$$

and $\hat{\mathbf{Q}} \equiv \frac{1}{n}\sum_{t=1}^{T}\mathbf{P}(\mathbf{X_t})\mathbf{P}(\mathbf{X_t})^T$. Then the functional estimator for $E(Z|\mathbf{X})$ is $h(\cdot) = \mathbf{p}(\cdot)^T\boldsymbol{\beta}$. Hence $\mu_t = Z_t - h(\mathbf{X_t})$ is the regression error. Define the long-run variance covariance matrix $(m \times m)$ for $\sqrt{\frac{1}{n}}\sum_{t=1}^{T}\mu_\mathbf{t}\mathbf{P}(\mathbf{X_t})$:

$$\mathbf{A} \equiv Var(\sqrt{\frac{1}{n}}\sum_{t=1}^{T}\mu_\mathbf{t}\mathbf{P}(\mathbf{X_t})).$$

Out-of-sample forecasts contain auto-correlation and heteroskedasticity of unknown form, and for inference in such model it is therefore necessary to consider the heteroskedasticity and auto-correlation consistent (HAC) estimator, which can consistently estimate the co-variance of model parameters. With HAC estimator $\hat{\mathbf{A}}$, the variance covariance matrix of the coefficients can be estimated with

$$\hat{\boldsymbol{\Omega}} \equiv \hat{\mathbf{Q}}^{-1}\hat{\mathbf{A}}\hat{\mathbf{Q}}^{-1}.$$

Therefore, the standard deviation of $\sqrt{\frac{1}{n}}(\hat{h}(x) - h(x))$ can be inferred as

$$\hat{\sigma}(x) \equiv \sqrt{\mathbf{P}(x)^T\hat{\boldsymbol{\Omega}}\mathbf{P}(x)}.$$

I can use the following implementation procedure to conduct the test:

1. First, simulate a random vector $\xi \sim \mathcal{N}(0, \hat{\boldsymbol{\Omega}})$ and let $\hat{t}(x) = \frac{\mathbf{P}(x)^T\xi}{\hat{\sigma}(x)}$.

2. Then simulate a large sample of ξ from step 1. Set $\bar{K}_n$ to be the $\bar{k}$-quantile of $\sup_{x\in\mathcal{F}}\hat{t}(x)$ in the sample, where $\bar{k} \equiv 1 - \frac{0.1}{\log T}$. Set

$$\hat{\Theta} \equiv \{x : \hat{h}(x) \geq \sup_{x\in\mathcal{F}}[\hat{h}(x) - \sqrt{\frac{1}{n}}\bar{K}_n\hat{\sigma}(x)] - 2\sqrt{\frac{1}{n}}\bar{K}_n\hat{\sigma}(x)\}.$$

[6] $P(x)$ contains the constant function $(p_1(x) = 1)$.

3. Third, let $\bar{k}_{1-\alpha}$ be the $(1-\alpha)$ quantile of $\sup_{x\in\hat{\Theta}} \hat{t}(x)$, and the $(1-\alpha)$ lower confidence bound $\hat{\eta}_{1-\alpha}$ is

$$\hat{\eta}_{1-\alpha} = \sup_{x\in\mathcal{F}} [\hat{h}(x) - \sqrt{\frac{1}{n}}\hat{k}_{1-\alpha}\hat{\sigma}(x)].$$

Reject the null hypothesis with $(1-\alpha)$ significance lower bound if $\hat{\eta}_{1-\alpha} > 0$.

Figure 2.5–Figure 2.7 illustrate CSPA tests of the weighted directional forecasts of a variety of machine learning models with respect to zero-profit random walk. Solid curves are fitted values of conditional weighted directional forecasts, and dashed curves are their corresponding lower confidence bounds. Note that the width of the confidnece bounds is affected by the probability mass of the conditional variable, with relatively tighter bounds around the median inflation rate differential, and wider bounds around the tails that imply historically rare scenarios. Solid curves in the figures show that all of the models showm more positive conditional expected profit than the random walk benchmark in some conditioning states, as $\hat{h}$ is above zero. I reject the CSPA hypothesis for the random walk: No model significantly outperforms the random walk at any value of the conditional variable. However, no model is absolutely more dorminant than the rest as the estimated conditional moment functions interect. Take the result for the Japan/U.S., for example: The ridge model performs better than the random walk during deflation and low growth episodes in Japan. However, we see from the unconditional weighted directional test result in Table 2.6 that the ridge forecasts fail to reject the null hypothesis.

Figure 2.2–Figure 2.4 show the joint CSPA test with five machine learning models as alternatives. Each lightly colored curve corresponds to the conditional moment function, and the black solid curve and dashed confidence bound concern the CSPA joint hypothesis. In particlar, Figure 2.2 shows that the null hypothesis is rejected for any value of the conditional states, as the confidence bound is always positive. In Figure 2.3 and Figure 2.4, although the maximum of the conditional moment function is positive for the majority of states of the inflation differential, the lower confidence bound is not always significantly positive. Also, note that the conditional expected function of the joint test achieves the highest on the tails of the conditional variable distribution. This implies that the machine learning models

demonstrate higher profitibiilty during less likely episodes of the inflation differential. For the robustness check, I also consider the GDP growth rate differential as the conditioning variable and arrive a similar result.

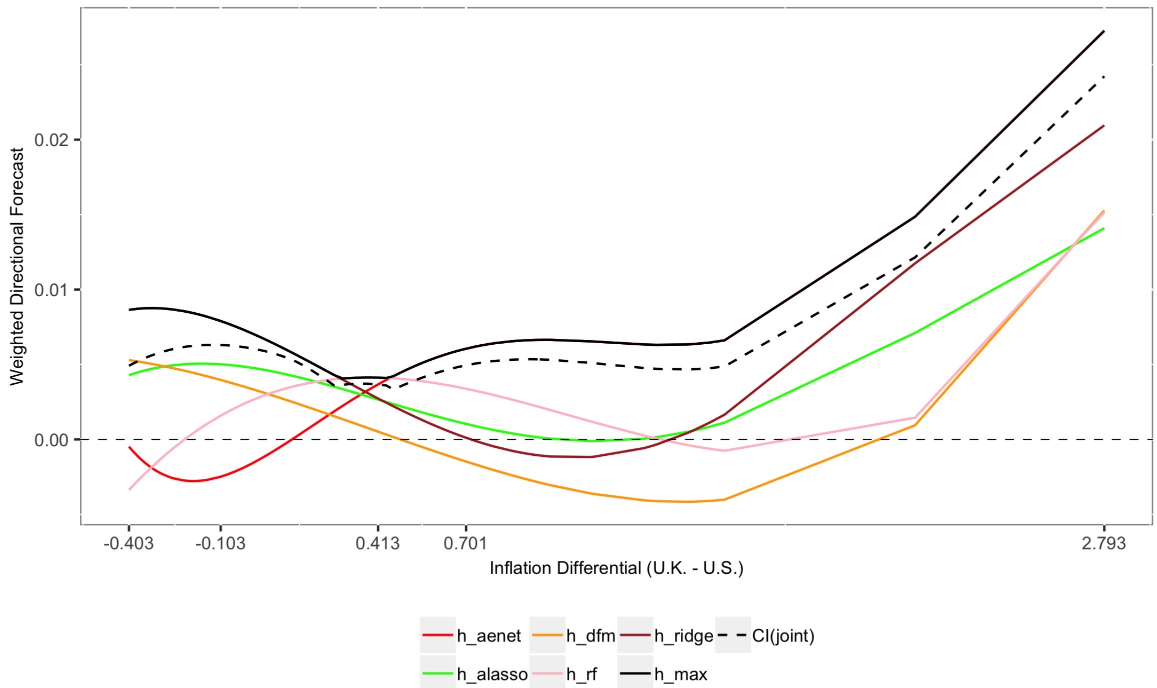


Figure 2.2: Weighted Directional Forecasts: CSPA (U.S./U.K.)

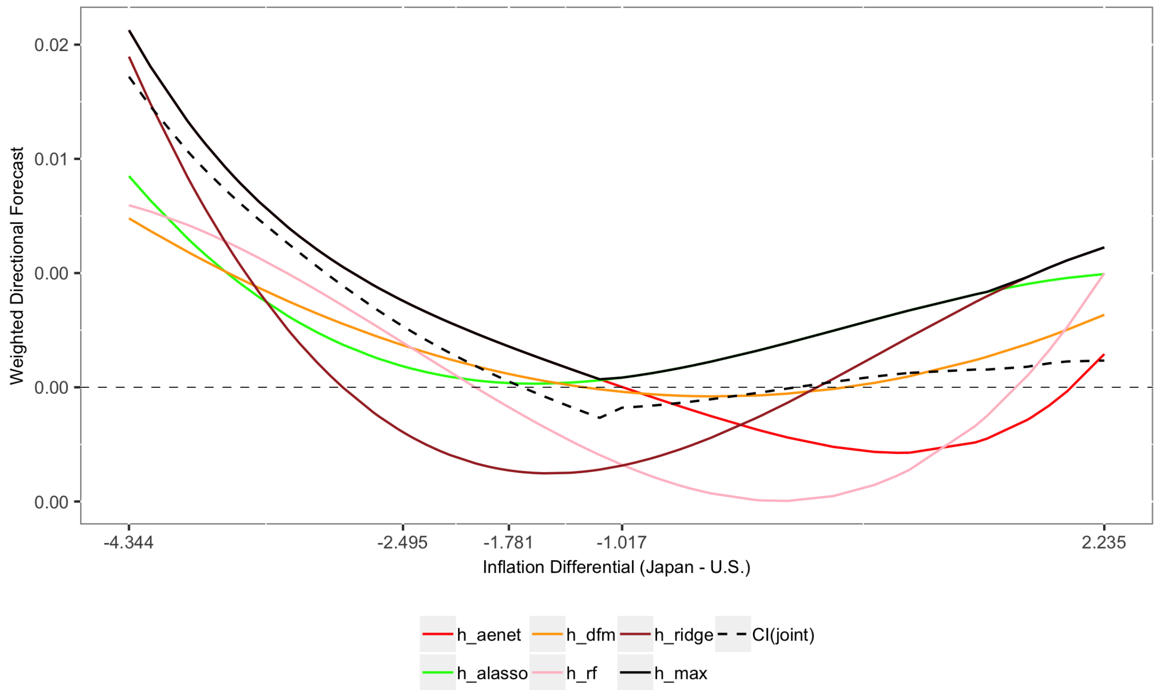


Figure 2.3: Weighted Directional Forecasts: CSPA (Japan/U.S.)

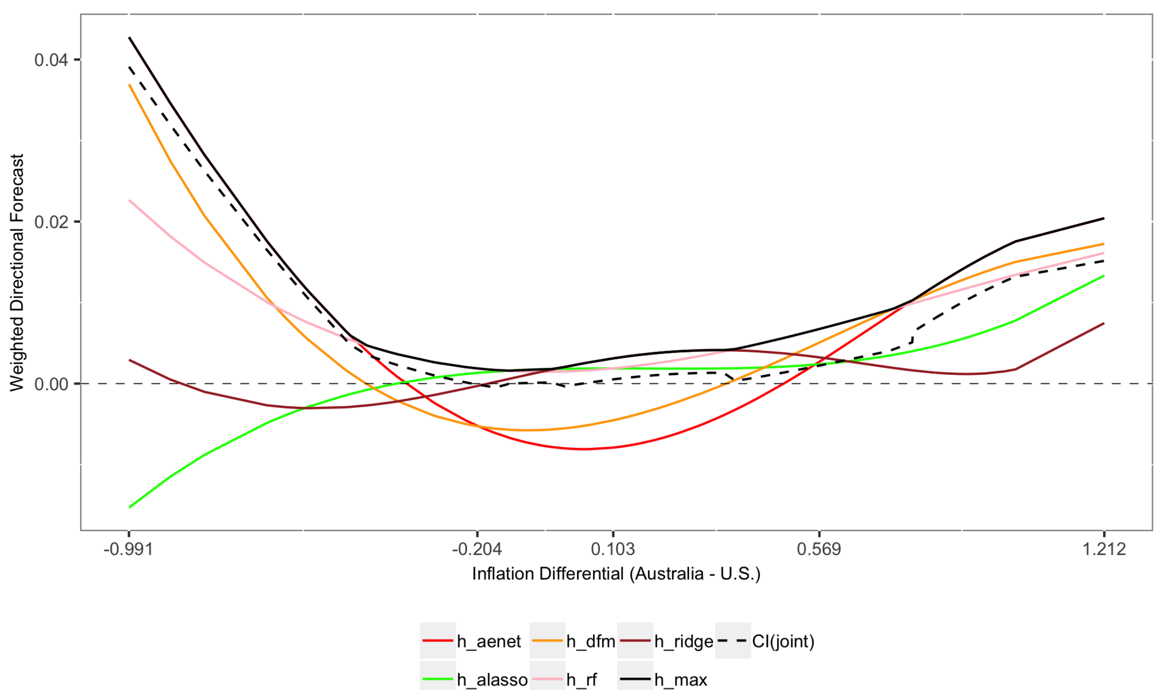


Figure 2.4: Weighted Directional Forecasts: CSPA (U.S./Australia)

2.6.3 Robustness Check

In this subsection, I analyze the robustness of the unconditional and conditional weighted directional performance results in four dimensions. First, I try different rolling-window sizes and Table 2.9 and Table 2.8 report MSE ratios for point forecasts and success ratios for directional forecasts. Second, I consider partitioning the sample periods into two subperiods: 2009M3–2014M2 and 2014M3–2019M1 for the directional forecasts. I report the success ratios and average weighted directional tests in Tables 2.10–2.11. Third, I consider the directional forecasts with different forecasting horizon $h = 3, 6$, and 12. Lastly, I consider a different conditioning variable, the GDP growth rate differential, to evaluate the conditional weighted directional tests.

1. **Diffferent Sizes of Rolling-Window Forecasting Scheme.** Table 2.8 shows the MSE ratios of point forecasts of the machine learning models for different rolling-window sizes. Table 2.9 shows the success ratios that correspond to different rolling-window sizes. It shows that the majority of MSE ratios are smaller than zero in the range of four- to six-year estimation windows. The 60-month rolling window achieves the largest average success ratio among different machine learning models compared with other estimation sizes.

2. **One-month-ahead Directional Forecasts in Two Subperiods.** Table 2.10 shows the success ratios of directional forecasts if we partition the sample into two subperiods, and Table 2.11 shows the result of the weighted directional test. As we can see, the forecast success ratios for the first half-period are higher than the second half for three currencies. Also, the weighted directional tests are less significant in the second half-period.

3. **Directional Forecasts with Forecasting Horizon $h = 3, 6$, and 12.** Table 2.12–Table 2.14 report the results of success ratios with forcasting horizons $h = 3, 6$, and 12 for three currencies. It is shown that the longer the forecasting horizon, the lower the success ratio of the directional forecast. This suggests that the Taylor rule-related Google Trends data are useful for forecasting short-term exchange rate movements.

4. **Conditional Weighted Directional Test (Conditioning Variable: GDP Growth differential).** Figure 2.8 to Figure 2.13 plot weighted directional forecasts conditional on the GDP growth differential of two countries. It shows similar results as in the case of the inflation rate differential. These figures reinforce the finding that the machine learning models are more profitable during less likely episodes of the conditioning variable.

2.7 Conclusion

This chapter studies the forecastability of exchange rate movements with Taylor rule-related Google Trends data. I evaluate the performance of three types of forecasts. The DMW test of the point forecasts of at least one machine learing model rejects the null hypothesis of equal forecasting errors equal to those of the random walk without drift for three currencies. According to the weighted directional test, the test statistics show significantly better performance of my models than the benchmark at the 10% significance level for all currencies. Furthermore, the conditional weighted directional forecasts allow us to know exactly *when* my models are more profitbale than the random walk with zero profit. I find that my weighted directional forecasts are significantly positive, especially on the tails of the conditioning variable. Also, the robustness results show that Google Trends data can help forecast exchange rate movement on a short horizon.

Table 2.7: Taylor Rule Model: Replication of Molodtsova and Papell (2009)

Currency	MSE (Taylor rule model)	MSE (random walk)	MSE ratio	DMW test statistic ($p-$value)
U.S./U.K.	$1.43e^{-3}$	$7.80e^{-4}$	0.547	$2.188**$ ($1.006e^{-2}$)
Japan/U.S.	$1.39e^{-3}$	$9.17e^{-4}$	0.656	$2.103**$ ($1.986e^{-2}$)
U.S./Australia	$1.53e^{-3}$	$8.00e^{-4}$	0.522	$2.895***$ ($2.651e^{-3}$)

Note: This table shows MSEs for both Taylor rule forecasts and random walk model forecasts for three currency pairs. It also reports the DMW test statistic and its respective p-value. $*$, $**$, $***$ indicate that the alternative model significantly outperforms the random walk at 10%, 5%, and 1% significance levels, respectively, based on standard normal critical values for the one-sided test. Forecasts results based on out-of-sample forecasts are from 2009M3 to 2019M1 for all currency pairs.

2.8 Appendix

Table 2.12: Directional Forecasts Success Ratios for U.S./U.K. ($h = 3, 6$, and 12)

Model	h=3	h=6	h=12
Adaptive Elastic Net	0.611	0.682	0.695
Random Forest	0.504	0.364	0.421
Adaptive LASSO	0.531	0.477	0.442
Ridge	0.549	0.551	0.505
Dynamic Factor Model	0.611	0.682	0.693

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts for two subperiods.

Table 2.8: Point Forecasts MSE Ratios For Different Rolling Window Sizes

	Rolling Window Size: $w = 48$		
Model	U.S./U.K.	Japan/U.S.	U.S./Australia
Adaptive Elastic Net	0.859	0.816	0.854
Random Forest	0.905	0.720	0.818
Adaptive LASSO	0.870	0.870	0.880
Ridge	0.958	0.893	0.985
Dynamic Factor Model	0.896	1.062	0.940
	Rolling Window Size: $w = 60$		
Adaptive Elastic Net	0.828	0.727	0.756
Random Forest	0.705	0.609	0.571
Adaptive LASSO	0.839	0.779	0.780
Ridge	1.039	0.783	0.785
Dynamic Factor Model	0.631	0.986	1.296
	Rolling Window Size: $w = 72$		
Adaptive Elastic Net	0.889	0.710	0.751
Random Forest	0.923	0.722	0.675
Adaptive LASSO	0.930	0.871	0.698
Ridge	1.009	0.892	0.875
Dynamic Factor Model	0.723	0.900	1.080

Note: This table shows MSE ratios of five machine learning forecasts over random walk model forecasts for three currencies with different rolling-window sizes.

Table 2.9: Directional Forecast Success Ratios for Different Rolling-Window Sizes

Model	U.S./U.K.	Japan/U.S.	U.S./Australia
	Rolling-Window Size: $w = 48$		
Adaptive Elastic Net	0.569	0.562	0.577
Random Forest	0.454	0.523	0.554
Adaptive LASSO	0.431	0.515	0.561
Ridge	0.432	0.538	0.577
Dynamic Factor Model	0.462	0.515	0.508
	Rolling-Window Size: $w = 60$		
Adaptive Elastic Net	0.602	0.639	0.529
Random Forest	0.576	0.588	0.580
Adaptive LASSO	0.542	0.521	0.597
Ridge	0.585	0.521	0.580
Dynamic Factor Model	0.534	0.546	0.563
	Rolling-Window Size: $w = 72$		
Adaptive Elastic Net	0.632	0.585	0.623
Random Forest	0.575	0.519	0.547
Adaptive LASSO	0.509	0.519	0.566
Ridge	0.491	0.509	0.462
Dynamic Factor Model	0.491	0.594	0.566

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts. Out-of-sample forecasts are from 2009M3 to 2019M1.

Table 2.10: Directional Forecasts Success Ratios for Two Subperiods

Model	U.S./U.K.	Japan/U.S.	U.S./Australia
	First half-period: 2009M3-2014M2		
Adaptive Elastic Net	0.644	0.610	0.593
Random Forest	0.644	0.576	0.576
Adaptive LASSO	0.576	0.559	0.542
Ridge	0.610	0.559	0.576
Dynamic Factor Model	0.576	0.525	0.525
	Second half-period: 2014M3–2019M1		
Adaptive Elastic Net	0.559	0.661	0.593
Random Forest	0.508	0.600	0.583
Adaptive LASSO	0.508	0.483	0.617
Ridge	0.559	0.483	0.583
Dynamic Factor Model	0.492	0.567	0.600

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts for two subperiods.

Table 2.11: Unconditional Weighted Directional Forecast Test Statistic for Two Subperiods

Model	U.S./U.K.	Japan/U.S.	U.S./Australia
	First half-period: 2009M3-2014M2		
Adaptive Elastic Net	$1.950**$	0.893	$1.600*$
	(0.025)	(0.186)	(0.055)
Random Forest	$1.990**$	-0.454	$2.070**$
	(0.023)	(0.675)	(0.0192)
Adaptive LASSO	$1.480*$	0.261	0.225
	(0.068)	(0.397)	(0.411)
Ridge	$1.710**$	-0.233	0.367
	(0.043)	(0.592)	(0.357)
Dynamic Factor Model	1.160	$1.400*$	0.351
	(0.123)	(0.081)	(0.363)
	Second half-period: 2014M3–2019M1		
Adaptive Elastic Net	0.272	0.814	0.233
	(0.393)	(0.208)	(0.408)
Random Forest	0.061	$1.390*$	0.834
	(0.476)	(0.082)	(0.202)
Adaptive LASSO	0.660	0.991	0.955
	(0.255)	(0.161)	(0.170)
Ridge	1.21	0.338	1.000
	(0.113)	(0.368)	(0.158)
Dynamic Factor Model	0.238	0.419	1.200
	(0.406)	(0.338)	(0.115)

Note: This table shows t-statistics for weighted directional forecasts for each currency and each machine learning model. $**$, $**$ and $*$ represent the 1%, 5%, and 10% significance levels, respectively. Out-of-sample forecasts are from 2009M3 to 2019M1.

Table 2.13: Directional Forecasts Success Ratios for Japan/U.S. ($h = 3, 6$, and 12)

Model	h=3	h=6	h=12
Adaptive Elastic Net	0.693	0.694	0.604
Random Forest	0.711	0.556	0.469
Adaptive LASSO	0.518	0.463	0.521
Ridge	0.553	0.519	0.500
Dynamic Factor Model	0.711	0.630	0.573

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts for two subperiods.

Table 2.14: Directional Forecasts Success Ratios for U.S./Australia ($h = 3, 6$, and 12)

Model	h=3	h=6	h=12
Adaptive Elastic Net	0.623	0.694	0.646
Random Forest	0.544	0.583	0.500
Adaptive LASSO	0.605	0.500	0.448
Ridge	0.505	0.469	0.448
Dynamic Factor Model	0.588	0.593	0.573

Note: This table shows directional forecast success ratios. Success ratios are calculated as the number of correct forecasts divided by the total number of out-of-sample forecasts for two subperiods.

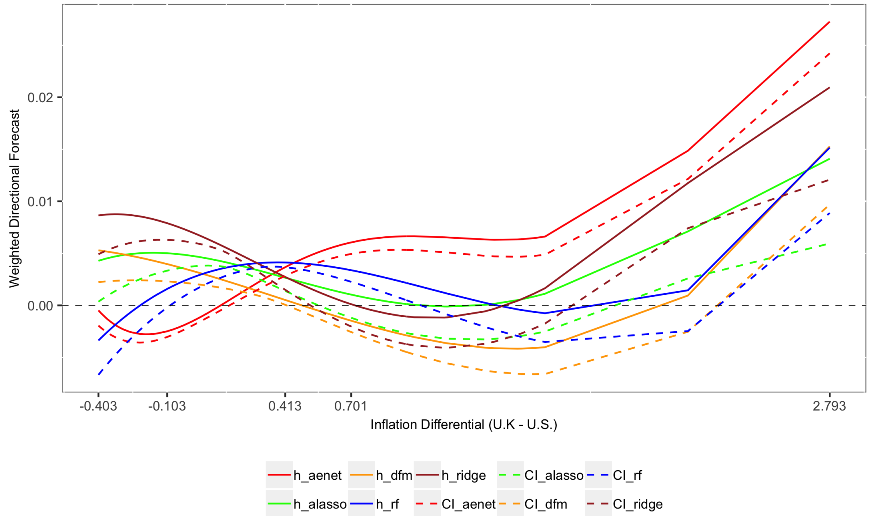


Figure 2.5: Weighted Directional Forecasts: CSPA (all, U.S./U.K.)

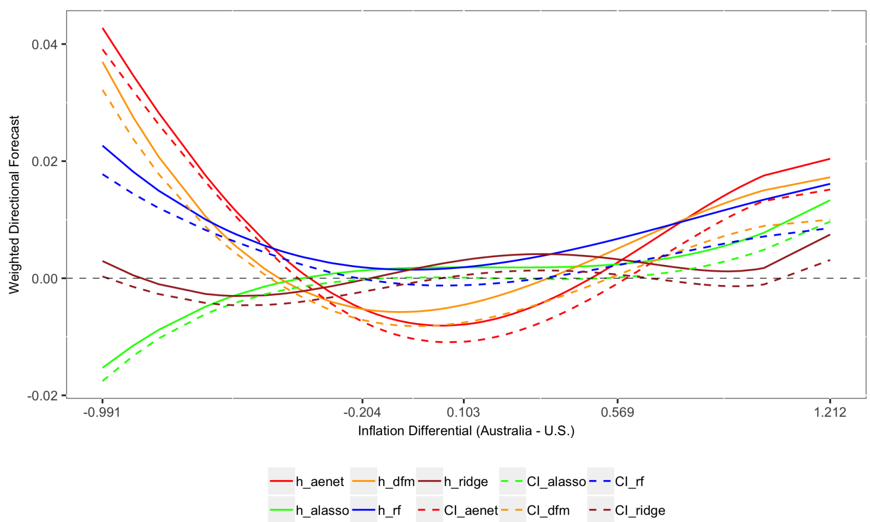


Figure 2.6: Weighted Directional Forecasts: CSPA (all, Japan/U.S.)

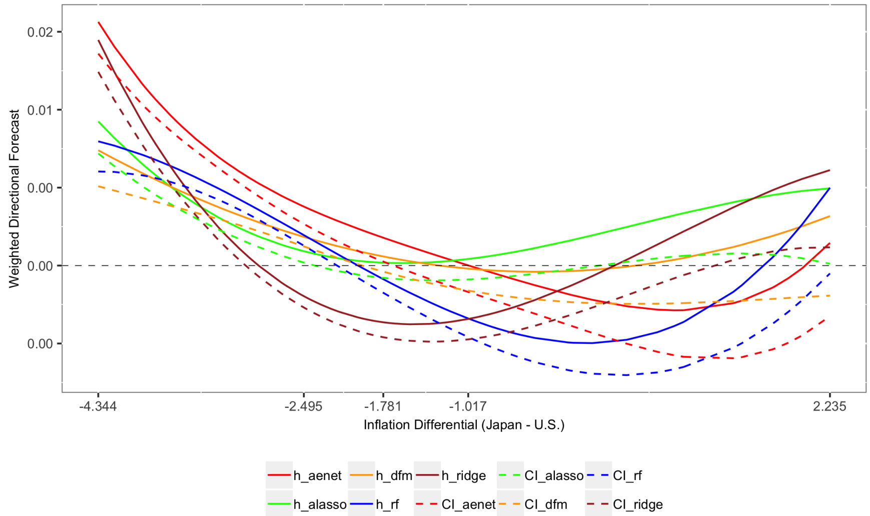


Figure 2.7: Weighted Directional Forecasts: CSPA (all, U.S./Australia)

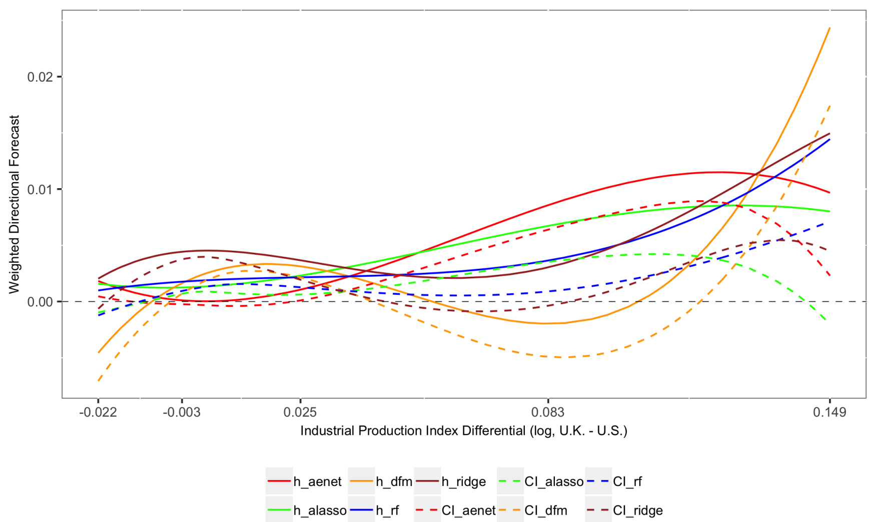


Figure 2.8: Weighted Directional Forecasts (IP): CSPA (all, U.S./U.K.)

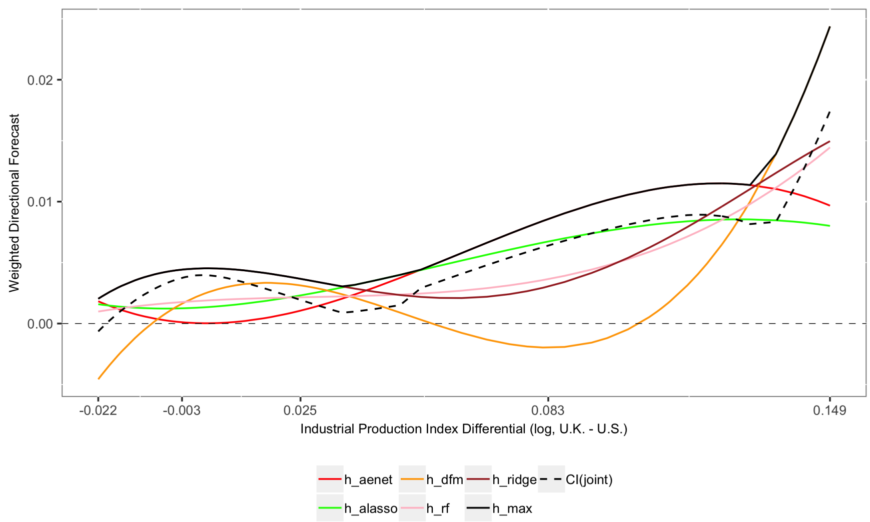


Figure 2.9: Weighted Directional Forecasts (IP): CSPA (U.S./U.K.)

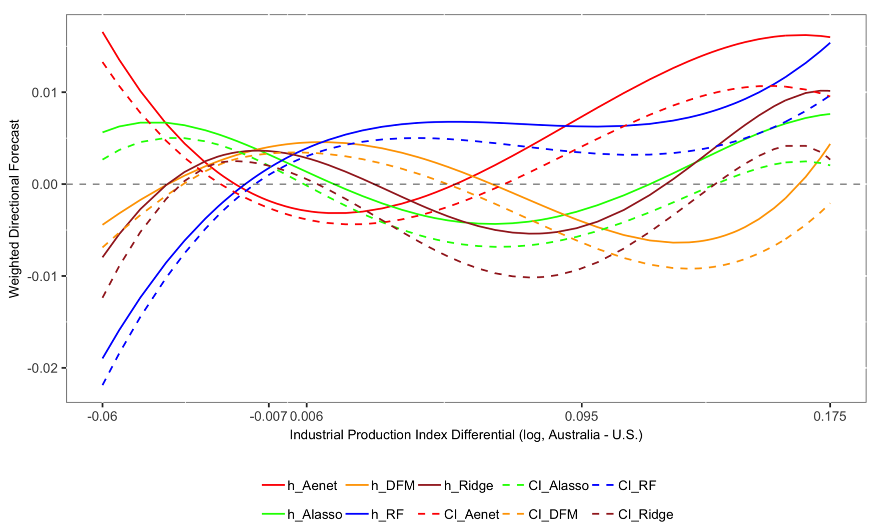


Figure 2.10: Weighted Directional Forecasts (IP): CSPA (all, U.S./Australia)

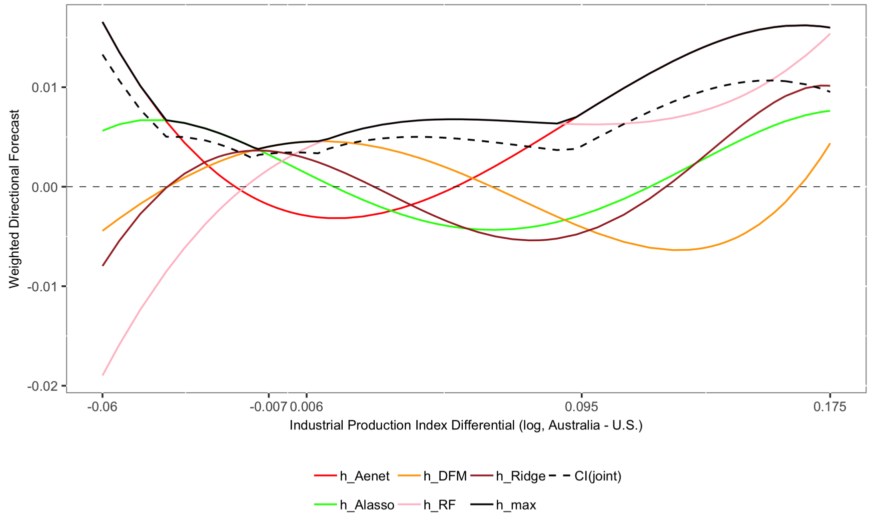


Figure 2.11: Weighted Directional Forecasts (IP): CSPA (U.S./Australia)

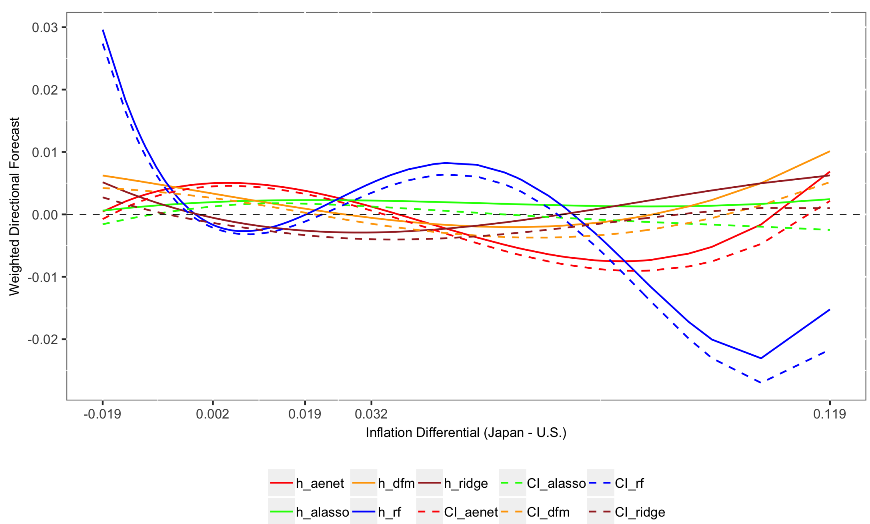


Figure 2.12: Weighted Directional Forecasts (IP): CSPA (all, Japan/U.S.)

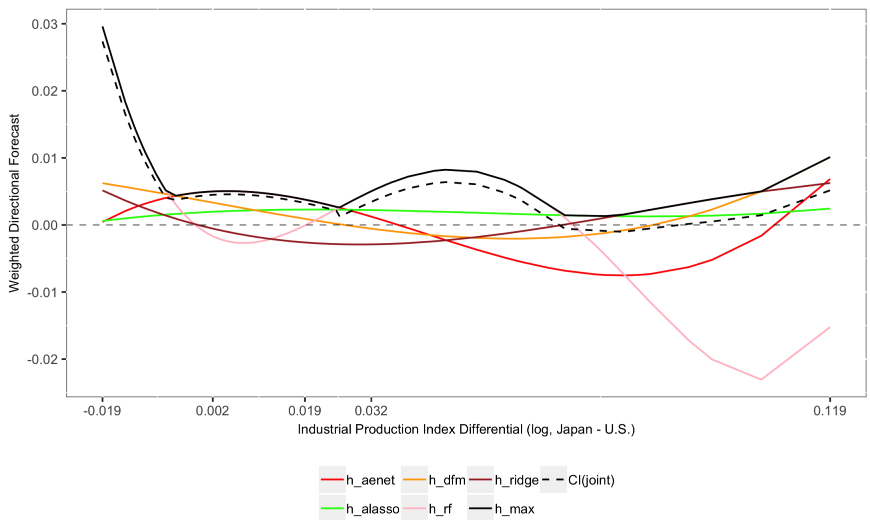


Figure 2.13: Weighted Directional Forecasts (IP): CSPA (Japan/U.S.)

CHAPTER 3

Interpreting Authorities' Reception of Fund Policy Advice: A Topic-Based Sentiment Analysis

3.1 Introduction

Traction is an important attribute of IMF policy advice, and gaps in traction are a risk for the Fund. Traction has several aspects, such as whether authorities "follow" Fund advice, whether authorities seek out Fund advice, and whether Fund advice is received positively or negatively by authorities. This paper focuses on the latter aspect. We use natural language processing tools and machine learning models to examine Article IV consultations for the period 2012-18 and establish some stylized facts for the tone and sentiment with which authorities receive Fund advice. The analysis focuses on monetary and fiscal policy but can be extended to other aspects of Fund advice. The methodology may also be useful for other Fund purposes.

Specifically, the approach entails the application of a latent Dirichlet allocation (LDA) model to Article IV staff reports. The model is able to recognize distinct policy topics and the words most commonly associated with each. After being "taught" how to read the reports' writing style, the model is able to predict the positive, negative, or neutral tones of authorities' responses to Fund advice. The results can be grouped according to sutopics within policy areas, thereby providing a granular picture of how Fund advice is received, as well as according to country groups. The stylized facts can provide useful input for assessing the impact of Fund advice, particularly if the approach is extended to other policy areas.

We adopt a topic-based approach to determine the sentiment tone of the authorities' views that are recorded in Article IV staff reports. Since each policy discussion consists of a range of subtopics, we apply the LDA model to objectively classify each paragraph in the Policy Discussion and each sentence in the Authorities' Views section into distinct policy themes. Specifically, the algorithm classifies monetary policy into five topics (monetary policy stance, exchange rate policy, macro-prudential policy, monetary policy framework, and financial sector supervision) and fiscal policy into six topics (tax policy, fiscal policy stance, expenditure policy, financing needs, medium-term fiscal policy framework, and impact of fiscal policy on real economy).

Simultaneously, we construct the topic-specific sentiment score, which is a product of the sentiment of the sentence-level authorities' views and the estimated probability of a specific topic of the sentence from the LDA model. The sentiment is defined as whether Fund advice is received positively or negatively by countries' authorities. The sentiment is predicted by the machine learning models as negative (-1), neutral (0), or positive (+1). We consider a variety of supervised machine learning models, including multinomial logistic regression, decision trees, random forest, Adaboost, etc. to predict the sentence in the Authorities' Views in Article IV reports after being provideed with a "gold standard" training set that has been labeled and validated by experienced economists. As a result, the model achieves around an 85% prediction accuracy rate in the testing set.

This allows us to measure multi-aspect policy sentiment by constructing the metrics to interpret authorities' tones in response to specific policy advice. For example, among all of the member countries, our results show that monetary policy stance sentiment is the lowest among all member countries and among other policy discussions over the period from 2012 to 2018. Yet tax policy, especially tax administration policy, achieves the highest sentiment score over those years. Moreover, across country groups, sentiment tones in both monetary and fiscal policy are relatively higher among low-income countries and emerging markets.

Our goal is to improve understanding of IMF policy advice as it is perceived by its member countries. This study is the first to use a topic-based sentiment analysis approach to Article IV consultations. Recently, there has been an explosion of studies in finance and economics using text as an alternative data source. In finance, textual data from news, social media, and company SEC filings are used to predict asset price movements (Tetlock, 2007; Engelberg and Parsons, 2011). In macroeconomics, textual data from FOMCs and news is proven to be predictive for macroeconomics data, and useful for estimating the effects of policy uncertainty (Scott and Varian, 2014; Baker, Bloom and Davis, 2013; Jegadeesh and Wu, 2015). In political economy, congressional text is used to study the dynamics of political debate (Gentzkow and Shapiro, 2010). In this paper, we apply textual analysis to IMF Article IV consultations and extract country authorities' sentiment in response to the policy advice proposed by the Fund.

The rest of the paper is organized as follows. Section 3.2 describes our sample and data sources. Section 3.3 introduces the topic-based sentiment analysis methodology. Section 4 reports the multi-aspect policy sentiment metrics, and Section 5 concludes.

3.2 Data

The IMF is an organization of 189 countries, and its mandate is to offer its member countries advice on economic policies that affect members' domestic and external stability in the form of Article IV staff reports. This paper focuses on historical Article IV consultations[1] from 2012 to 2018 received by all of its member countries.

Each Article IV consultation follows a structured writing style. They consist of several major sections, such as Recent Developments, Economic Outlook and Risks and Policy Discussions. Specifically, during an Article IV consultation, an IMF team of economists visits a member country to assess and discuss its economic and financial conditions and policies, such as monetary policy, fiscal policy, financial sector policy, etc. The assessments and policy recommendations are recorded in the Policy Discussion section. Moreover, following the Policy Discussion, the counties' authorities express their commitment to or disagreement with the proposed policy recommendations in the Authorities' Views section. For the purpose of this paper, we focus our attention on monetary and fiscal policy only.

[1]The reports are publicly available online at https://www.imf.org/external/index.htm.

We obtain balanced coverage of Article IV staff reports by country groups from 2012 to 2018. Our sample consists of around 657 Article IV consultations. For each report, we locate and scrape the content in the Policy Discussion section using a web crawler and regular expression tools. Then we conduct the following parsing procedure: (1) we separate fiscal policy and monetary policy content from the Policy Discussion by using regular expressions; (2) for each policy area, we separate Authorities' Views content from the policy recommendations and; (3) we break the policy recommendation into paragraphs and Authorities' Views into sentences[2]. Throughout this procedure, we successfully produce around 2,000 paragraphs for monetary policy and fiscal policy discussions. Meanwhile, we obtain around 1,200 and 1,400 sentences for the Authorities' Views in its corresponding policy area.

3.3 Methodology

In this section, we lay out the construction of our multi-aspect topic-based sentiment measurements. There are several challenges in analyzing these multifaceted documents, because each policy discussion (either monetary or fiscal) is a mixture of a range of subtopics. For example, monetary policy can contain policy assessments of and recommendations for monetary policy stance, exchange rate policy, macro-prudential policy, monetary policy framework, and financial sector supervision. In Section 3.3.1, we introduce a topic modeling technique—latent Dirichlet allocation (LDA) model—to automatically summarize and isolate the content of each subtopic prior to sentiment analysis. Another complication arises from the sentiment analysis. There are two broad approaches to conducting the sentiment analysis: (1) apply a bag-of-words model first to filter the corpus down to only words that are thought to carry sentiment and use a predefined dictionary of positive and negative words to calculate the average sentiment score and; (2) generate uni-/bi-/skip-gram models from the tokenized text and apply classification models to train the machine to truly "understand" the language it uses in the text. Because the language style in our sample is distinct from other

[2]We do so because each paragraph in the policy recommendation tends to focus on one topic, and authorities' views usually only have one paragraph, in which each sentence represents one piece of opinion about a particular policy theme.

data sources, such as news and social media, we use a more robust and objective predictive method to conduct the sentiment analysis. Section 3.3.2 introduces the models.

3.3.1 Topic Modeling

First, the goal is to classify all sentences in the Authorities' Views by their underlying policy theme. We classify each sentence from the Authorities' Views into distinct policy topics with the LDA algorithm which is, a probabilistic topic model developed by Blei et al. (2003). The LDA model allows sets of observed sample units, which are called documents (D), to be explained by latent structures. It is an unsupervised learning clustering algorithm, and we don't need to label the documents before the sample is classified. The model assumes that each document (a paragraph or sentence) can be summarized as a distribution over a set of distinct topics, and each topic is a distribution of the words that appear in all documents. When observing the sample documents, we can estimate the underlying topic structure that the documents are generated from. Both of the unobserved latent distributions are assumed to belong to the Dirichlet family. The task is to estimate the posterior document and topic distributions using a Bayesian approach. We illustrate the LDA model with a relevant simple example. Suppose we are given three sentences:

1. *The stance of monetary policy was maintained accommodative.*
2. *Fiscal consolidation is key to maintain confidence in debt sustainability.*
3. *Monetary policy should stand ready to offset the contractionary effects of fiscal consolidation.*

After using some natural language processing tools, such as phrases groupings, tokenization, lemmatization, and stop words elimination, we construct the sample as the input for the LDA model:

1. *stance, monetary policy, maintain, accommodative*
2. *fiscal consolidation, key, maintain, confidence, debt, sustainability*

3. *monetary policy, stand, ready, offset, contractionary, effect, fiscal consolidation*

We can intuitively tell that the first sentence is about monetary policy, the second is about fiscal policy, and the third one is a mixture of both. If the LDA model performs well, it should output the posterior distribution of the topic in the documents by simply reading the ranked words/phrases:

$$
\begin{aligned}
&\hat{\beta}_1 \equiv \{\hat{P}_{topic1}(\textit{fiscal consolidation}), \hat{P}_{topic1}(maintain), \hat{P}_{topic1}(confidence), \\
&\quad \hat{P}_{topic1}(sustainability), ...\} \\
&\quad = \{0.124, 0.082, .0.081.0.077\} \\
&\hat{\beta}_1 \equiv \{\hat{P}_{topic2}(\textit{monetary policy}), \hat{P}_{topic2}(maintain), \hat{P}_{topic2}(accomodative), \hat{P}_{topic2}(stance), ...\} \\
&\quad = \{0.122, 0.119, .0.113.0.111\}
\end{aligned}
$$

Then we can name the first topic as monetary policy and the second as fiscal policy. The posterior topic mixture in each sentence corresponds to our intuition:

$$
\begin{aligned}
\hat{\theta}_1 &\equiv \{\hat{P}_{sentence1}(topic1), \hat{P}_{sentence1}(topic2)\} = \{0.169, 0.831\} \\
\hat{\theta}_2 &\equiv \{\hat{P}_{sentence2}(topic1), \hat{P}_{sentence2}(topic2)\} = \{0.903, 0.097\} \\
\hat{\theta}_3 &\equiv \{\hat{P}_{sentence3}(topic1), \hat{P}_{sentence3}(topic2)\} = \{0.301, 0.699\}.
\end{aligned}
$$

This says that the first sentence has a 16.9% probability of monetary policy and 83.1% of fiscal policy. We can similarly interpret for the second and third sentences.

From the above simple example, one can use the LDA model to successfully classify the input into distinct themes without manually labeling a large set of documents.

Before proceeding with the LDA classification of our sample, we need to specify the number of topics, N, that we want to identify for each policy area. To start, we set the range of the number of topics from 3 to 20. To get robust results, we use a model coherence measure based on point-wise mutual information (PMI) to compare different choices of topics. According to Blei et al. (2003), the LDA model assumes the following generative process for each document d in the sample space:

1. We denote the proportion of topic $n \in N$ in document $d \in \mathcal{D}$ as $\theta_{d,n}$, and assume the vector of topic proportions in document d, θ_d from a Dirichlet distribution with order N, parameter α:

$$\theta_d \sim Dir(\alpha).$$

2. We assign each word w in document d with topic $T_{d,w} \in [1, ..., N]$, and it follows a multinomial distribution conditional on document d's topic proportion vector θ_d. Namely, $T_{d,w} \sim \text{Multinomial}(\theta_d)$.

3. As in the previous simple example, the topic distribution matrix is denoted as $\beta \in \{\beta_1, ..., \beta_N\}$. Each topic distribution β_n follows a Dirichlet distribution that contains W words. Namely, $\beta_n \sim Dir(\mu)$.

4. Therefore, a word is drawn from a multinomial distribution conditional on topic T_d distribution, $P(w|T_d, \beta)$.

However, since this latent procedure is not observed by the human reader, we need to estimate the posterior distribution of this topic structure given the collection of words w in the documents using the Bayesian method. First, we write the joint distribution of a topic mixture θ, topics T, and words W given the parameters α and β:

$$P(\beta, \theta, \mathbf{T}, \mathbf{W}) = \prod_{n=1}^{N} P(\beta_n) \prod_{d=1}^{D} P(\theta) \prod_{w}^{I_d} (P(T_{d,w}|\theta_d) P(W_{d,w}|\beta, T_{d,w}).$$

By integrating over θ and T, the marginal distribution of document d is:

$$P(\mathbf{W}|\alpha, \beta) = \int p(\theta|\alpha) (\prod_{n=1}^{N} \sum_{T_n} P(T_n|\theta) P(W_n|T_n, \beta)) d\theta.$$

Therefore, we will obtain the posterior distribution of the document-topic latent structure given the observed words in Article IV reports:

$$P(\beta, \theta, \mathbf{T}|\mathbf{W}) = \frac{P(\beta, \theta, \mathbf{T}, \mathbf{W})}{P(\mathbf{W})}.$$

Blei et al. (2003) use a variational EM algorithm to estimate the parameters.

In order to obtain the optimal number of topics for each policy area, a coherence measurement, based on the PMI, can help to quantify the hanging and fitting together of information pieces. The PMI between any two words, i and j, is defined as

$$PMI(w_i, w_j) = \ln(\frac{P(w_i, w_j)}{P(w_i)P(w_j)}).$$

For example, assume there are 100 words in the sample documents. The word "monetary" appears 10 times, "policy" 20 times, and the frequency of the co-occurrences of both words is 8. Then the PMI is calculated as

$$PMI(\text{'monetary'}, \text{'policy'}) = \ln(\frac{\frac{8}{100}}{\frac{10}{100} \times \frac{20}{100}}) = 1.38.$$

Therefore, the coherence measure of the whole sample is

$$\text{Cohenrence} = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \ln(\frac{P(w_i, w_j)}{P(w_j)}).$$

A higher coherence measure indicates better topic classification results. Figure 3.1 plots the coherence scores for different numbers of topics[3] for both monetary policy and fiscal policy. It turns out that, there are 7 topics "hidden" in the monetary policy discussion and 6 in fiscal policy.

[3]Because monetary policy/fiscal policy in each Article IV consultation contains at least three subtopics, so we start from $N \geq 3$. However, when the number of topics increases, some topics become redundant and can lead to semantically less meaningful topics.

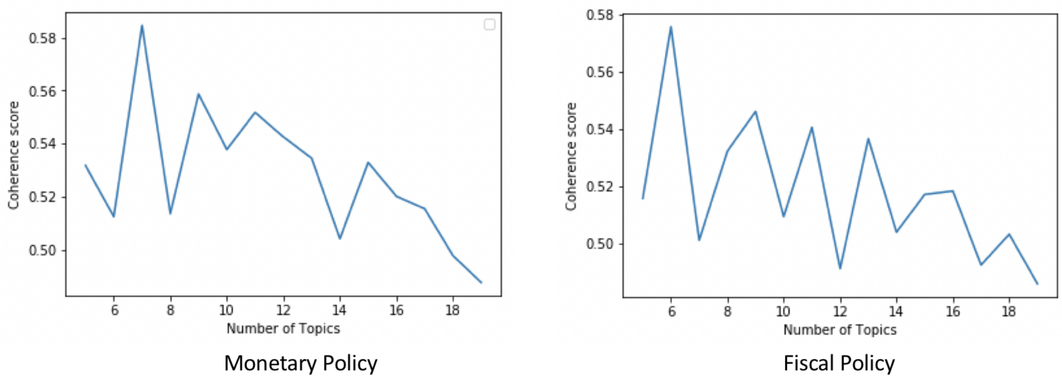


Figure 3.1: Coherence Measure for Different Number of Topics

Given the optimal number of topics, N, we can compute the posterior expectations of word distribution for each topics, $\{\hat{\beta}_1, ..., \hat{\beta}_N\}$, and posterior topic mixture, $\hat{\theta}_d$, for each document d in our sample. Table 3.2 and Table 3.3 report the top 30 words for each topic in monetary policy and fiscal policy, respectively. From the tables, we can clearly identify the topics according to these key words. For example, the first topic in Table 3.2 is about monetary policy stance, which consists of words such as *inflation*, *monetary policy*, *target*, and so on. The LDA model identifies themes in fiscal policy with almost no ambiguity. For example, Topic 3 in Table 3.3 contains keywords such as *spend*, *social*, *public*, *expenditure*, etc., suggesting expenditure policy. To sum up, the model identifies seven topics for monetary policy: monetary policy stance, monetary easing post-GFC, exchange rate/reserve policy, monetary policy framework, macro-prudential policy, financial sector supervision, and other topics[4]. The six themes in fiscal policy are tax policy[5], fiscal policy stance, expenditure policy, financing needs, medium-term fiscal policy framework, and impact of fiscal policy on real economy.

Besides the posterior word distribution for distinct topics, the model also outputs the posterior topic distribution, $\hat{\theta}_d = [\hat{\theta}_{d,1}, ..., \hat{\theta}_{d,N}]$, for each document, $d \in D$. In other words, $\hat{\theta}_{d,n}$ represents the proportion of topic n in document d.

[4]We are not able to clearly identify this group with the word distribution, so we categorize it as the other topics.

[5]Tax policy includes tax revenue policy as well as tax administration policy.

For more robust results, we select 100 documents for each policy area and ask our economists to read and manually check whether the topic distributions make sense in the documents. And ninety out of 100 documents are approved.

Monetary Policy Stance	Monetary Easing Post-GFC	Exchange Rate/Reserve Policy	Monetary Policy Framework	Macro-prudential Policy	Financial Sector Supervision	Other topics
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7
Words	Words	Words	Words	Words	Words	Words
inflation	policy	exchange_rate	monetary_policy	bank	financial	fiscal
monetary_policy	financial	reserve	framework	loan	bank	gdp
pressure	monetary	external	liquidity	risk	credit	public
remain	risk	market	bank	capital	stability	government
target	monetary_policy	flexibility	market	foreign	financial_sector	finance
increase	credit	fx	strengthen	asset	strengthen	debt
tighten	ease	intervention	improve	lend	include	increase
growth	interest_rate	volatility	operation	deposit	plan	budget
interest_rate	stability	international	monetary	limit	prudential	spend
monetary	condition	level	instrument	ratio	development	investment
year	measure	maintain	interest_rate	provision	risk	reform
expect	low	currency	policy	npls	access	sector
policy_rate	term	shock	include	high	regulation	deficit
inflation_expectations	lend	allow	government	balance_sheet	improve	revenue
price	credit_growth	import	use	requirement	aml_cft	tax
policy	rat	peg	enhance	monitor	address	growth
lower	economic	regime	develop	credit	sector	structural
current	economy	continue	excess	fund	reform	need
rate	negative	depreciation	transmission	large	supervision	support
risk	growth	current_account	new	household	law	project
low	macroprudential	policy	need	system	finance	private
need	demand	position	term	increase	system	consolidation
continue	need	limit	target	exposure	action	infrastructure
monetary_policy_stance	give	real	deposit	financial	effort	term
term	domestic	term	help	supervision	measure	medium
rise	limit	help	make	commercial	enhance	expect
stance	remain	give	mechanism	level	progress	measure
output	see	broadly	reduce	two	new	expenditure
inflationary	private	need	liquidity_management	remain	make	cost
demand	tighten	line	objective	measure	fund	external

Figure 3.2: Top 30 Words for Different Topics in Monetary Policy

Tax policy	Fiscal policy stance	Expenditure policy	Financing needs	Medium-term fiscal policy framework	Impact of fiscal policy on real economy
Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
Words	Words	Words	Words	Words	Words
tax	gdp	spend	finance	fiscal	growth
revenue	deficit	social	bank	policy	remain
reform	debt	public	debt	term	economy
improve	revenue	expenditure	government	medium	economic
include	budget	increase	domestic	rule	inflation
strengthen	year	government	risk	need	demand
public	target	wage	borrow	framework	export
administration	fiscal	pension	external	consolidation	external
vat	adjustment	reform	end	budget	expect
management	spend	investment	loan	support	decline
plan	project	infrastructure	liability	sustainability	risk
budget	expenditure	service	fund	risk	investment
system	measure	budget	billion	structural	year
need	expect	system	market	reform	increase
investment	public	need	arrears	debt	sector
effort	increase	reduce	public	growth	low
new	balance	benefit	credit	commitment	global
financial	primary	bill	project	balance	current_account
measure	ratio	capital	remain	target	real
income	growth	measure	private	long	crisis
increase	medium_term	include	million	buffer	lower
enhance	consolidation	cost	contingent	help	domestic
reduce	decline	age	foreign	give	rise
exemption	level	sector	cost	provide	private
subsidy	structural	save	continue	adjustment	exchange_rate
implement	lower	program	guarantee	space	import
government	reduce	health	financial	ensure	strong
price	government	plan	asset	consider	euro
expenditure	surplus	current	bond	shock	price
collection	percentage_point	improve	deposit	anchor	oil

Figure 3.3: Top 30 Words for Different Topics in Fiscal Policy

3.3.2 Sentiment Analysis

In order to construct the multi-aspect sentiment measure for the documents, in addition to topic distribution estimation we have obtained in Section 3.3.1, we need to know the overall sentiment tone of each sentence. In this section, we introduce a variety of machine learning models to predict the sentiment tone as *positive*, *negative*, or *neutral*.

First, we follow a bag-of-words approach to extract features from the textual data and build the regressor space. The approach simply calculates the frequency of each word in each document. With the previous example, we have:

1. *stance, monetary policy, maintain, accommodative*;
2. *fiscal consolidation, key, maintain, confidence, debt, sustainability*;
3. *monetary policy, stand, ready, offset, contractionary, effect, fiscal consolidation.*

Then for each document, we count the frequency of tokens from the dictionary {*"stance". "monetary policy", "maintain", "accommodative", "fiscal consolidation", "key", "confidence", "debt", "sustainability", "stand", "ready", "offeset, "contractionary", "effect"*}. For example, we can convert the first sentence in the example into a vector only containing "0" and "1": $[1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$. Similarly, we proceed and transform all sentences into vectors with only frequency counts.

Having prepared our regressor space, we then consider a variety of machine learning models for sentiment prediction.

Multinomial Logistic Regression. Multinomial logistic regression is used to predict categorical responses based on multiple independent variables. Independent variables are the word frequency vectors precessed by the bag-of-words method. Multinomial logistic regression is a simple extension of binary logistic regression that allows for more than two categories of dependent variables; in our case, there are three classes: *positive, negative*, and *neutral*. Also, multinomial logistic regression uses maximum likelihood estimation to evaluate the probability of categorical responses.

Decision Tree. A decision tree is a commonly used machine learning method for classification tasks based on high-dimensional regressor space. This method classifies a sentence into branch-like segments that construct tree with a root node, internal nodes, and terminal nodes. The model is nonparametric and is able to deal with high-dimensional datasets without imposing a parametric structure. Typically, A training dataset is used to build a decision tree and the optimal final model is determined using a validation dataset.

Random Forest. Random forest uses the concept of bagging and bootstrapping and is an ensemble learning method for classification tasks by combining multiple decision trees. Because dealing with a single decision tree will lead to the problem of overfitting, which may affect the overall classification result on the test set, random forest is much more robust and prone to noises. (1) Random forest uses the concept of bootstrapping, in which each decision tree is trained on the subset of the total training data and; (2) the results of all decision trees will be averaged and the result will be consistent (bagging).

Ridge Classifer. Ridge regression is a shrinkage technique for analyzing a high-dimensional regressor space that suffers from multicollinearity. Although least squares estimates are unbiased, their variances are large. By adding a regularization term to the coefficient estimates, ridge regression reduces prediction errors.

Adaboost. Adaboost is a boosting technique that combines multiple "weak classifiers" into a "strong classifier." This technique can be applied to any classification algorithm and has several advantages. First, it helps choose the training set for a new classifier to be trained based on the results of previous models. Also, it determines the weights of the classifiers when combining the results. Based on these benefits, we apply Adaboost to our multinomial logistic regression, random forest, and ridge regression.

Training, Test set, and Evaluation. As sentiment prediction is a difficult task, and since domain knowledge of Article IV writing style is necessary for annotating the overall sentiment tone, it is critical that our machine learning models are well trained before undertaking this task of predicting.

Our sample consists of around 2,800 sentences in the Authorities' Views in 657 Article IV consultations. We ask our economists to label the sentiments of 350 sentences as our **gold labels**; some of the annotated sentences are included in the appendix. Also, we split these labeled sentences into training (70%), validation (15%) and test (15%) sets. Specifically, the training set provides the "gold standard" that the machine learning models learn from. Models and features are evaluated and selected by examining out-of-sample prediction accuracy in the validation set, and the test set serves as the final performance of a selected model and important predictors.

The metric we use to evaluate the performance of the machine learning models on the test set is the prediction accuracy rate. Consider the contingency table shown in Figure 3.4. Each cell labels a set of possible outcomes. In our sentiment prediction case, for example, true positives are the documents that are labeled correctly (indicated by the gold labels)[6].

[6]Either positive, negative, or neutral

Therefore, the prediction accuracy rate is defined as

$$\text{Prediction Accuracy Rate} = \frac{\text{True positive} + \text{True negative} + \text{True neutral}}{350}.$$

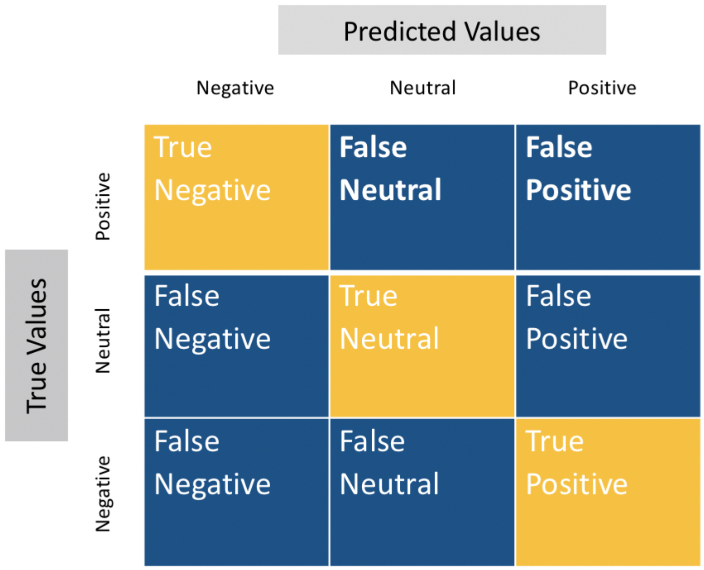


Figure 3.4: Prediction Contingency Table

For an alternative robust performance measure, we consider F-1 score[7]. The reason is that in our sample, the classes are unbalanced; there are more positive labeled sentences than negative ones. If we put emphasis on finding weak policy recommendations, the F-1 measure which considers both precision and recall, rewards us for detecting sentences with different classes, even in an unbalanced sample.

In order to tune the parameters of the models, five fold cross-validation is used for each of the machine learning classification models, with the results averaged over the five folds. Figure 3.4 displays prediction performance results on the validation set of the labeled sample, showing that the decision tree model achieves the highest prediction accuracy rate of 85%, with 900 features (bi-grams). Then, with the best selected model, we calculate

[7] Details are included in Appendix 3.6.2.

the prediction accuracy rate for the test set, and it achieves 84.6%[8]. Figure 3.5 shows the prediction contingency table of the test set with the decision tree.

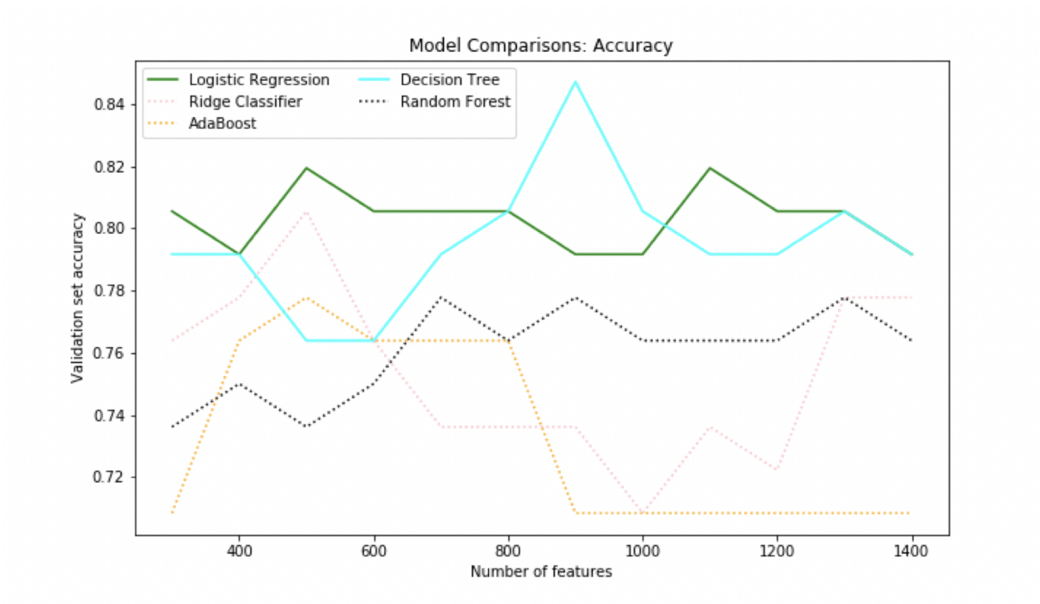


Figure 3.5: Machine Learning Prediction Performance

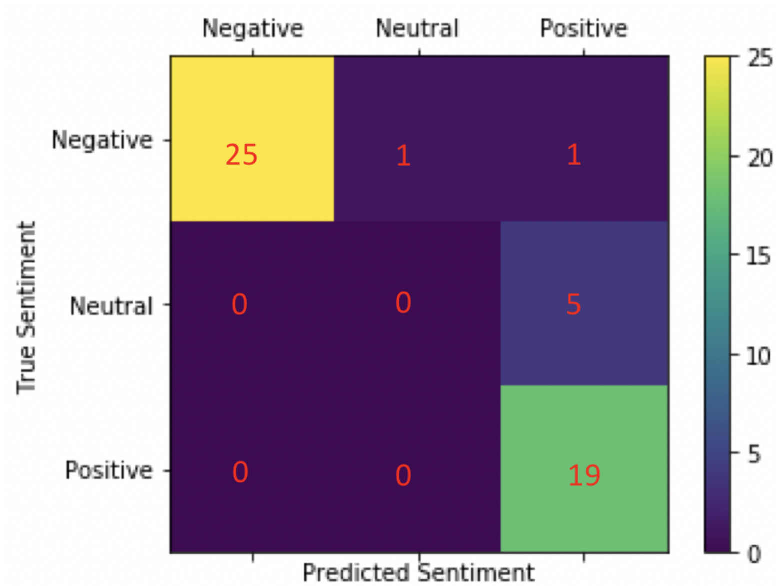


Figure 3.6: Prediction Contingency Table on the Test Set (52 Sentences)

[8]However, none of the neutral tones is predicted correctly.

3.3.3 Multi-aspect Topic-based Sentiment Metrics

After we obtain the topic distribution for each document in our sample as illustrated in Section 3.3.1 and overall predicted sentiment as illustrated in Section 3.3.2, we can define the multi-aspect sentiment score, $Tone$. Since there are six[9] policy themes in monetary policy and six for fiscal policy, the topic contents consist of N-level sentiment scores. Explicitly, for sentence d in country a, year t Article IV consultation, with the estimated topic proportion $\hat{\theta}_n$ and the sentiment tone $\hat{S}$, we define the topic-based sentiment $Tone$ as

$$Tone_{d,n}^{a,t} = \hat{\theta}_{d,n}^{a,t} \hat{S}_d^{a,t}.$$

A higher (lower) $Tone$ means a more positive (negative) sentiment. Figure 3.7 demonstrates an example of the construction of a multi-aspect sentiment score for document, d.

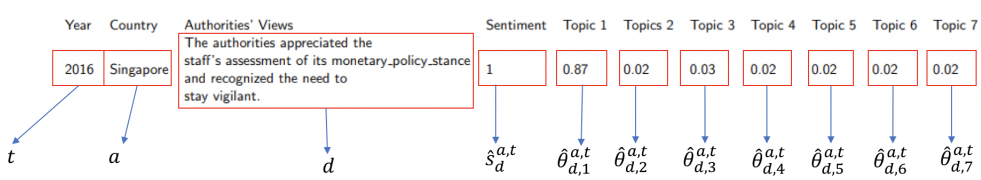


Figure 3.7: An Example of Multi-aspect Sentiment Score

Next, we are interested in aggregating topics-based $Tone_{d,n}^{a,t}$ to each Article IV level for each country-year pair:

$$Tone_n^{a,t} = \sum_{d=1}^{D_{a,t}} Tone_{d,n}^{a,t},$$

where $D_{a,t}$ is the total number of sentences in the authorities' views section in year t for country a. Moreover, if we aggregate further arrive at the weighted topic-specific policy sentiment measure for different country groups. For example, for low-income countries (LICs), the topic n policy sentiment score is

$$Tone_n^{a,LICs} = \sum_{a \in LICs} Tone_n^{a,t} * \frac{1}{\sum_{a \in LICs} \sum_d I_d^{a,t}} * 100,$$

[9]We neglect the ambiguous "other topics."

where $I_d^{a,t}$ is a dummy that indicates whether document d belongs to a particular country group.

3.4 Results

In the context of monetary policy, we established some stylized facts of the aggregate sentiment measure for six subtopics among all member countries from 2012–2018. In Figure 3.8, the green bar represents scores that only count the positive sentiment, the red bar represents negative sentiments, and diamonds represent net sentiment scores. The dashed blue line is the average net sentiment across all policy themes.

For example, as a core mandate of the Fund, exchange rate policy receives higher net sentiment. For the monetary policy framework, authorities in general welcome Fund policy advice on how to improve their monetary framework in the short term and over the medium term, with some of them requesting the Fund's technical assistance (TA).

If we zoom in on each topic, in Figure 3.10 we can see how sentiment varies across different country groups. G7 and G20 express more dispersed sentiments. However, among emerging markets, low income countries, and fragile states, fund policy advice receives more agreement and commitment. Under the monetary policy stance topic, one reason for the more negative sentiment among G7 countries may because they have stronger analytical capacity and market participation, yet emerging markets and fragile states tend to more rely on the Fund's analysis.

Moreover, in the scope of fiscal policy, tax policy (including tax administration) has the highest sentiment among all member countries. Yet for other fiscal policy themes, sentiments tend to be balanced and authorities' receptions of them is mixed. Expenditure policy overall tends toward mixed. One reason may that advice on cutting social spending to increase fiscal suitability is sometimes pushed back on by country authorities.

Sentiment scores also vary among different country groups. G7 and advanced economies express a more negative tone toward the tax policy. In contrast, EM, LICs, and fragile states welcome or voice commitment to the tax policy. The fiscal policy framework tends to be positively received by all country groups.

3.5 Conclusion

Our paper presents a topic-based approach to determine the sentiment tone of the authorities' views on monetary and fiscal policy recorded in Article IV staff reports. Moreover, our analysis can be extended to study other aspects and policy areas of Fund advice, such as in financial sector and structural reforms. Also, in order to achieve a more accurate sentiment prediction, a larger training set is needed for classification models to learn from. This requires continuing efforts to extend "gold labeled" sentences. Furthermore, based on established stylized facts of topic-based sentiment, we can study the drivers behind the reception of different Fund policy advice by different country groups.

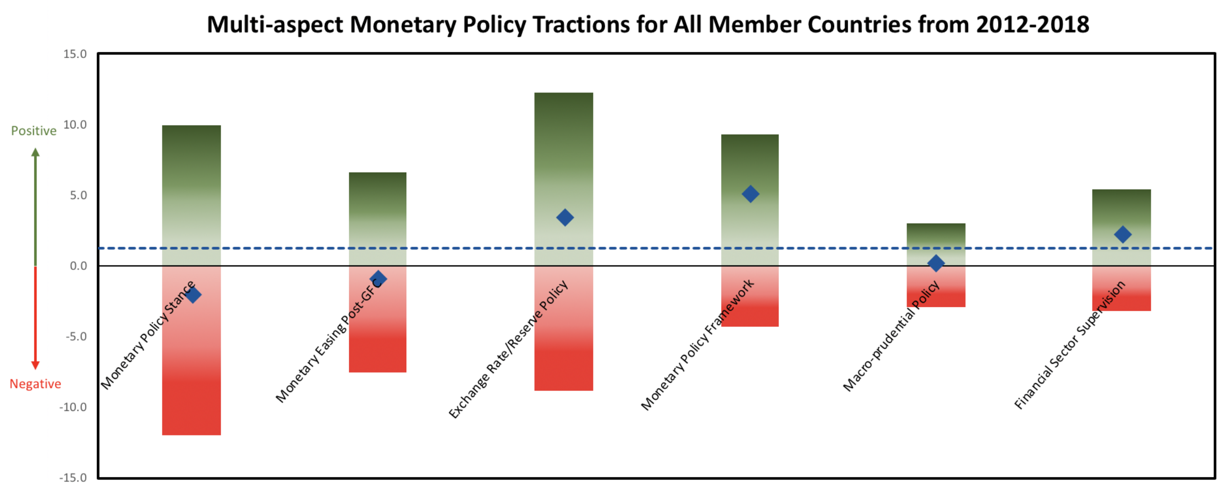


Figure 3.8: Multi-aspect Monetary Policy Tractions for All Members 2012-2018

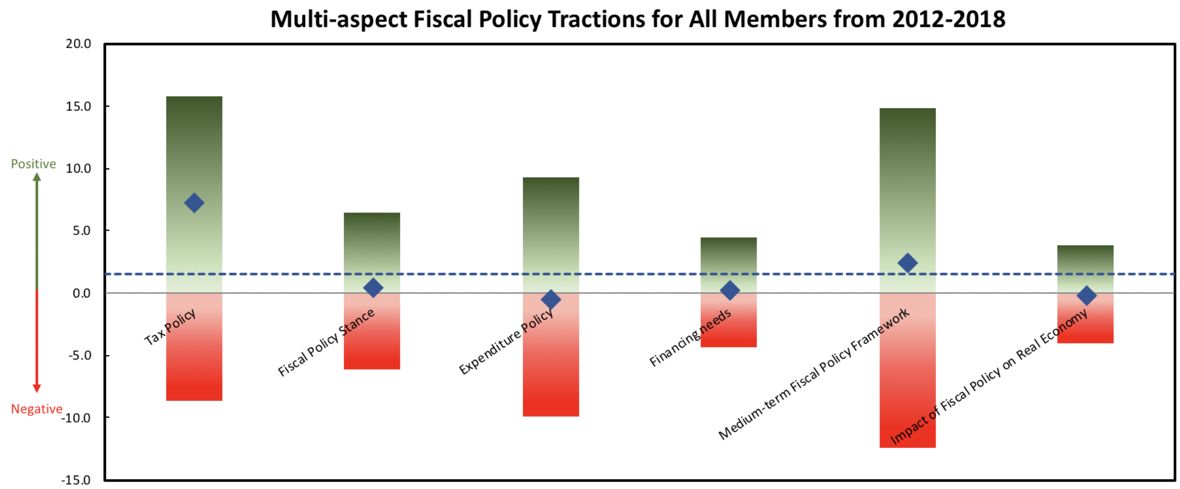


Figure 3.9: Multi-aspect Fiscal Policy Tractions for All Members 2012-2018

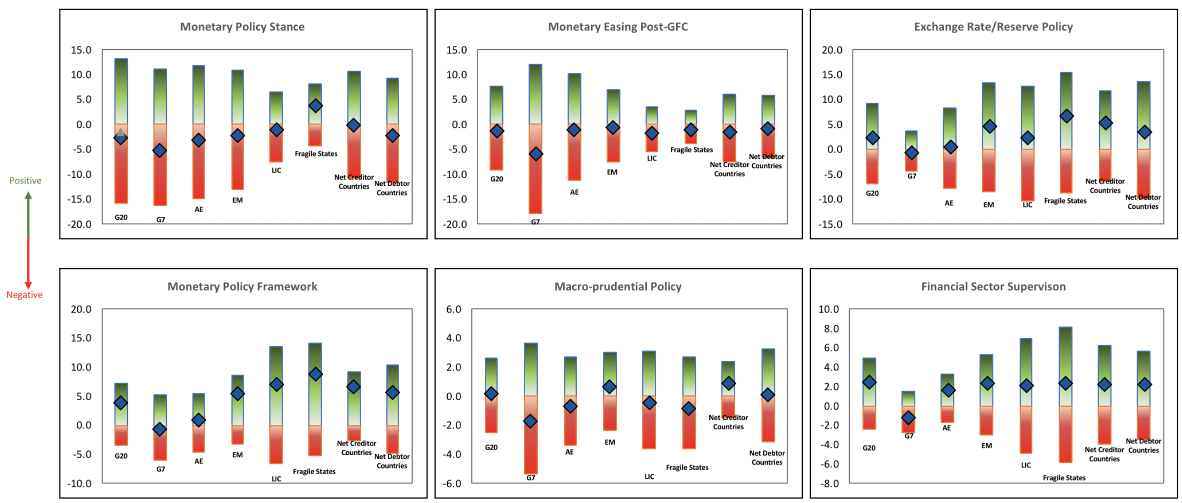


Figure 3.10: Multi-aspect Monetary Policy Tractions for Country Groups 2012-2018

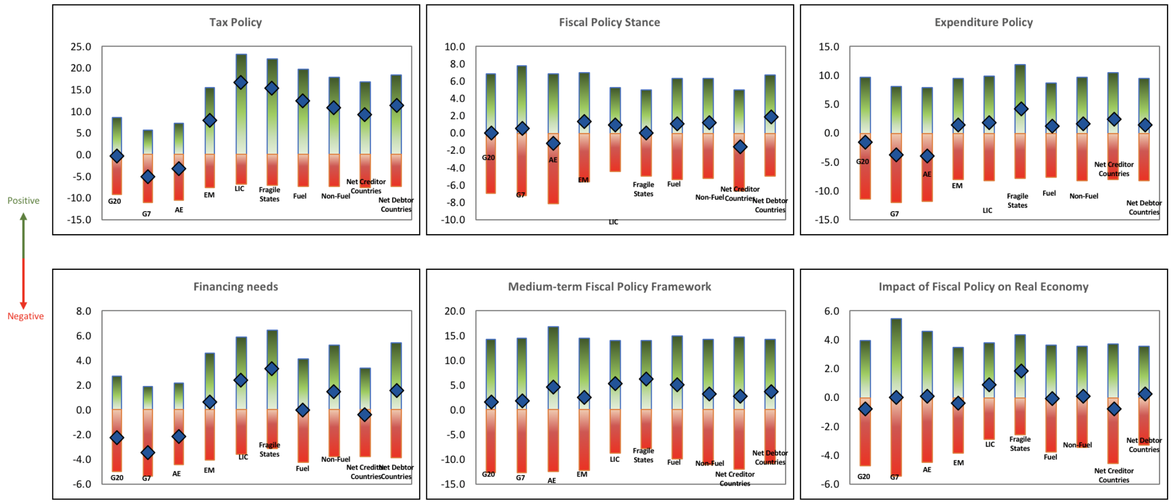


Figure 3.11: Multi-aspect Fiscal Policy Tractions for Country Groups 2012-2018

3.6 Appendix

3.6.1 Example: labeled Sentences

1. Negative tones:

   The authorities broadly agreed with the staff's recommendations, but noted that comprehensive reform takes time. (-1)

   The authorities agreed with the need to use exhaustible resources efficiently, but mentioned that the resource horizon is likely to increase as recent exploration activities point to new discoveries (e.g., in the natural gas sector). (-1)

   The authorities favored a looser monetary policy stance to promote growth. (-1)

2. Positive tones:

   The authorities also welcomed the ta they have been receiving for the transition of the national accounts to SNA93. (+1)

   The authorities welcomed continued Fund technical assistance. (+1)

   The authorities expressed their commitment to supporting the peg and building reserves over the medium term. (+1)

3.6.2 Prediction Performance Measure: F-1 Score

Refer to the contingency table in Table 3.4. Precision is the percentage of sentences that are predicted to be positive that are in fact positive. It is defined as

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positve}}.$$

In contrast, recall measures the percentage of sentences that are correctly predicted by the model. It is defined as

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}.$$

Therefore, the F-1 score is a weighted harmonic average of precision and recall. It is defined as

$$F - 1 = \frac{2\ \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

REFERENCES

[Allcott and Gentzkow, 2017] Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

[Aruoba et al., 2009] Aruoba, S. B., Diebold, F. X., and Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics*, 27(4):417–427.

[Askitas and Zimmermann, 2009] Askitas, N. and Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting.

[Atkeson et al., 2001] Atkeson, A., Ohanian, L. E., et al. (2001). Are phillips curves useful for forecasting inflation? *Federal Reserve bank of Minneapolis quarterly review*, 25(1):2–11.

[Bai and Ng, 2008] Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

[Baker et al., 2016] Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

[Bańbura et al., 2013] Bańbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting*, volume 2, pages 195–237. Elsevier.

[Bickel et al., 2009] Bickel, P. J., Ritov, Y., Tsybakov, A. B., et al. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732.

[Binici and Cheung, 2012] Binici, M. and Cheung, Y.-W. (2012). Exchange rate dynamics under alternative optimal interest rate rules. *Pacific-Basin Finance Journal*, 20(1):122–150.

[Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

[Bok et al., 2018] Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018). Macroeconomic nowcasting and forecasting with big data. *Annual Review of Economics*, 10:615–643.

[Brandl and Pickl, 2009] Brandl, B. and Pickl, S. (2009). Increasing the fitness of fundamental exchange rate forecast models. *Int. J. Contemp. Math. Sciences*, 4(16):779–798.

[Bulut, 2018] Bulut, L. (2018). Google trends and the forecasting performance of exchange rate models. *Journal of Forecasting*, 37(3):303–315.

[Carriero et al., 2015] Carriero, A., Clark, T. E., and Marcellino, M. (2015). Realtime nowcasting with a bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(4):837–862.

[Carriero et al., 2018] Carriero, A., Clark, T. E., and Marcellino, M. G. (2018). Assessing international commonality in macroeconomic uncertainty and its effects.

[Castelnuovo and Tran, 2017] Castelnuovo, E. and Tran, T. D. (2017). Google it up! a google trends-based uncertainty index for the united states and australia. *Economics Letters*, 161:149–153.

[Chernis and Sekkel, 2017] Chernis, T. and Sekkel, R. (2017). A dynamic factor model for nowcasting canadian gdp growth. *Empirical Economics*, 53(1):217–234.

[Chernozhukov et al., 2013] Chernozhukov, V., Lee, S., and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737.

[Chetverikov et al., 2016] Chetverikov, D., Liao, Z., and Chernozhukov, V. (2016). On cross-validatsed lasso. *arXiv preprint arXiv:1605.02214*.

[Choi and Varian, 2009] Choi, H. and Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, pages 1–5.

[Choi and Varian, 2012] Choi, H. and Varian, H. (2012). Predicting the present with google trends. *Economic Record*, 88:2–9.

[Chojnowski and Dybka, 2017] Chojnowski, M. and Dybka, P. (2017). Is exchange rate moody? estimating the influence of market sentiments with google trends. *Econometric Research in Finance*, 2(1):1–21.

[Clark and West, 2006] Clark, T. E. and West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis. *Journal of econometrics*, 135(1-2):155–186.

[Clark and West, 2007] Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311.

[Dahlhaus et al., 2017] Dahlhaus, T., Guénette, J.-D., and Vasishtha, G. (2017). Nowcasting bric+ m in real time. *International Journal of Forecasting*, 33(4):915–935.

[De Mol et al., 2008] De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

[Dunis and Williams, 2002] Dunis, C. and Williams, M. (2002). Modelling and trading the eur/usd exchange rate: Do neural network models perform better? *Derivatives use, trading and regulation*, 8(3):211–239.

[Efron et al., 2004] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

[Eichenbaum et al., 2017] Eichenbaum, M., Johannsen, B. K., and Rebelo, S. (2017). Monetary policy and the predictability of nominal exchange rates. Technical report, National Bureau of Economic Research.

[Engel et al., 2015] Engel, C., Mark, N. C., and West, K. D. (2015). Factor model forecasts of exchange rates. *Econometric Reviews*, 34(1-2):32–55.

[Engel et al., 2007] Engel, C., Mark, N. C., West, K. D., Rogoff, K., and Rossi, B. (2007). Exchange rate models are not as bad as you think [with comments and discussion]. *NBER macroeconomics annual*, 22:381–473.

[Engelberg and Parsons, 2011] Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1):67–97.

[Fan and Li, 2001] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, 96(456):1348–1360.

[Fan and Lv, 2008] Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911.

[Ferrara and Marsilli, 2014] Ferrara, L. and Marsilli, C. (2014). Nowcasting global economic growth: A factor-augmented mixed-frequency approach.

[Gentzkow et al., 2017] Gentzkow, M., Kelly, B. T., and Taddy, M. (2017). Text as data. Technical report, National Bureau of Economic Research.

[Gentzkow and Shapiro, 2010] Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

[Geweke, 1977] Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

[Giacomini and White, 2006] Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

[Giannone et al., 2018] Giannone, D., Monti, F., and Sokol, A. (2018). A new approach to nowcasting with mixed-frequency bayesian vars. Technical report, mimeo.

[Giannone et al., 2004] Giannone, D., Reichlin, L., and Sala, L. (2004). Monetary policy in real time. *NBER macroeconomics annual*, 19:161–200.

[Giannone et al., 2008] Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

[Ginsberg et al., 2009] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012.

[Goddard et al., 2015] Goddard, J., Kita, A., and Wang, Q. (2015). Investor attention and fx market volatility. *Journal of International Financial Markets, Institutions and Money*, 38:79–96.

[Gourinchas and Rey, 2007] Gourinchas, P.-O. and Rey, H. (2007). International financial adjustment. *Journal of political economy*, 115(4):665–703.

[Hamid and Heiden, 2015] Hamid, A. and Heiden, M. (2015). Forecasting volatility with empirical similarity and google trends. *Journal of Economic Behavior & Organization*, 117:62–81.

[Harvey et al., 1997] Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

[Hasselgren et al., 2018] Hasselgren, A., Peltomäki, J., and Graham, M. (2018). Speculator activity and cross-asset predictability of fx returns.

[Jegadeesh and Wu, 2017] Jegadeesh, N. and Wu, D. A. (2017). Deciphering fedspeak: The information content of fomc meetings.

[Kim et al., 2014] Kim, Y. J., Liao, Z., and Tornell, A. (2014). Speculators positions and exchange rate forecasts: Beating random walk models. *Unpublished Manuscript UCLA*.

[Koop and Onorante, 2013] Koop, G. and Onorante, L. (2013). Macroeconomic nowcasting using google probabilities. *University of Strathclyde*.

[Li and Liao, 2018] Li, J. and Liao, Z. (2018). Uniform nonparametric inference for time series.

[Li et al., 2019] Li, J., Liao, Z., and Quaedvlieg, R. (2019). Conditional superior predictive ability.

[Li et al., 2015] Li, X., Shang, W., Wang, S., and Ma, J. (2015). A midas modelling framework for chinese inflation index forecast incorporating google search data. *Electronic Commerce Research and Applications*, 14(2):112–125.

[Loughran and McDonald, 2011] Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

[Marcellino et al., 2018] Marcellino, M. G., Papailias, F., Mazzi, G. L., Kapetanios, G., and Buono, D. (2018). Big data econometrics: Now casting and early estimates.

[Markiewicz et al., 2018] Markiewicz, A., Verschoor, W. F., Zwinkels, R. C., et al. (2018). Inattentive search for currency fundamentals. *Willem FC and Zwinkels, Remco CJ, Inattentive Search for Currency Fundamentals (August 03, 2018)*.

[Meese and Rogoff, 1983] Meese, R. A. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics*, 14(1-2):3–24.

[Mohebbi et al., 2011] Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H., and Kumar, S. (2011). Google correlate whitepaper.

[Molodtsova and Papell, 2012] Molodtsova, T. and Papell, D. (2012). Taylor rule exchange rate forecasting during the financial crisis. Technical report, National Bureau of Economic Research.

[Molodtsova and Papell, 2009] Molodtsova, T. and Papell, D. H. (2009). Out-of-sample exchange rate predictability with taylor rule fundamentals. *Journal of international economics*, 77(2):167–180.

[Nicholson, 2015] Nicholson, W. (2015). Bigvar user's guide dimension reduction procedures for multivariate time series.

[Pennington et al., 2014] Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[Plakandaras et al., 2015] Plakandaras, V., Papadimitriou, T., Gogas, P., and Diamantaras, K. (2015). Market sentiment and exchange rate directional forecasting. *Algorithmic Finance*, 4(1-2):69–79.

[Pradeepkumar and Ravi, 2016] Pradeepkumar, D. and Ravi, V. (2016). Forex rate prediction using chaos and quantile regression random forest. In *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, pages 517–522. IEEE.

[Rossi, 2013] Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, 51(4):1063–1119.

[Scott and Varian, 2013] Scott, S. L. and Varian, H. R. (2013). Predicting the present with bayesian structural time series. *Available at SSRN 2304426*.

[Scotti, 2016] Scotti, C. (2016). Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises. *Journal of Monetary Economics*, 82:1–19.

[Stock and Watson, 2011] Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford handbook on economic forecasting*.

[Stock and Watson, 2005] Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research.

[Taylor, 1993] Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, volume 39, pages 195–214. Elsevier.

[Tetlock, 2007] Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

[Uematsu and Tanaka, 2017] Uematsu, Y. and Tanaka, S. (2017). Macroeconomic forecasting and variable selection with a very large number of predictors: A penalized regression approach.

[Varian, 2014] Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

[Vozlyublennaia, 2014] Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, 41:17–35.

[Wang and Wu, 2012] Wang, J. and Wu, J. J. (2012). The taylor rule and forecast intervals for exchange rates. *Journal of Money, Credit and banking*, 44(1):103–144.

[Zou, 2006] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

[Zou and Hastie, 2005] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

[Zou and Zhang, 2009] Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics*, 37(4):1733.