Application of Google Trends Data in Exchange Rate Prediction

Motoki Masuda and Fumiko Takeda[+]

Department of Technology Management for Innovation

University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 Japan

Tel: +81-3-5841-1191

Email: motoki.masuda@gmail.com (Masuda)

Email: takeda@tmi.t.u-tokyo.ac.jp (Takeda)

February 2019

Abstract

This study explores the possibilities of applying Google Trends to exchange rate forecasting. Specifically, we use Google Trends to capture market sentiment in Japan and the United States and construct a sentiment index. We forecast the one-month-ahead USD/JPY rates using three structural models and two autoregressive models and examine whether our sentiment index can improve the predictive power of these models. The data we use run from January 2004 to August 2018, treating January 2004 to February 2011 as the training sample and March 2011 to August 2018 as the forecast sample. We find that the addition of the sentiment index into the autoregressive models decreases the mean squared prediction error. We also test the sentiment indices of different word numbers and find that the 25- and 30-word indices perform best; in particular, the 30-word index improves all the models tested in this study.

Keywords: exchange rate; forecast; international finance; search frequency; search engine; yen-dollar rate

JEL classification: F31; F37; G15; G17;

[+] Author in correspondence.

1. Introduction

The importance of investor sentiment in explaining asset prices is becoming increasingly important (De Long et al. 1990; Da et al. 2014). With the development of the Internet, data from social networks, online articles, search engines, and so on are allowing us to reveal sentiment in ways we never could have a few decades ago. In this study, we use the publicly available tool Google Trends to help us reveal market sentiment in Japan and the United States and predict the exchange rate movements between these two countries.

Google Trends allows us to specify the period, country, and region in which we are interested and provides time series data of the search frequencies. These search frequencies are normalized to take values from zero to 100 (where the week/month with the highest search frequency is 100), and this value is known as the Search Volume Index (SVI). This value also accounts for any apparent increase in search volume due to an overall increase in Internet users.

Many studies have shown that data from Google Trends can be used to explain and predict certain real-life phenomena. Ginsberg et al. (2009) find that the SVI on influenza-related terms helps predict the number of flu-related physician visits two weeks before the official reports by the US Centers for Disease Control and Prevention. Choi and Varian (2012) find that the SVI helps predict unemployment claims and automobile sales. In the field of finance, Google Trends has also been used to explain stock price movements using the search frequencies of company names as a proxy of investor attention (Da et al. 2011; Bank et al. 2011; Adachi et al. 2017). Da et al. (2014) create a FEARS (Financial and Economic Attitudes Revealed by Search) index with economy-related search frequencies as a proxy of investor sentiment and use it to explain stock price movements and various other assets in the United States.

Following this line of research and Da et al. (2014), in this study we construct a sentiment index (SI) for Japan and the United States to capture market sentiment in each of these countries and verify the extent of its ability to explain one-month-ahead monthly USD/JPY rates. We use these indices as additional regressors to examine whether they improve the predictive power of three structural exchange rate models (i.e., the purchasing power parity (PPP) model, interest rate parity (IRP) model, and monetary model) and two simple autoregressive models (AR(1) and AR(2)). As measures of forecast accuracy, we use the mean squared prediction error (MSPE), Clark and West's (2007) test of equal predictive accuracy (CW test hereafter), and a direction of change (DOC) test as well as compare the performance of all the models against a benchmark random walk (RW) model. We use the monthly exchange rate and macroeconomic data from January 2004 to August 2018, splitting the data into an initial training sample (January 2004 to February 2011) and a forecast sample (February 2011 to August 2018).

Overall, we find that the SI does improve the predictive accuracy of the exchange rate models in various aspects. Our best results are seen with AR(1) and AR(2). These base models show a decrease in the MSPE of 6.09% and 6.99% compared with the RW model, and the addition of the SI improves these percentages to 6.38% and 8.65%, respectively. Furthermore, the former model's ability to predict the directional changes of exchange rate returns increases by a statistically significant amount. We also test the SIs of different word numbers (10, 20, 25, 30, 35, and 40) and find that the 25- and 30-word indices perform best; in particular, the 30-word index improves all the models tested in this study to a certain degree.

Exchange rate forecasting is a topic of interest not only for researchers, but also for practitioners such as investors and exporters/importers. Our contribution to the literature is twofold. First, aside from the U.S. SI, similar to Da et al. (2014), we construct a unique SI for Japan using Google Trends data and a Japanese dictionary. Second, we improve exchange rate models by including the SIs as independent variables. In particular, we show that this Google Trends SI has the power to help us predict one-month-ahead movements of the USD/JPY rate.

The rest of the paper is organized as follows. In Section 2, we review past studies and provide background information. In Section 3, we explain the data, models, and methods of our forecast analysis. We show and discuss our main results in Section 4, and in Section 5 we conduct additional analyses as a robustness check. Lastly, we conclude in Section 6.

2. Literature Review and Background Information

Exchange rate forecasting is a difficult task that has been tackled by many researchers over the years, with moderate success. There is no single correct way to calculate the "par" value of an exchange rate; this is what makes predicting exchange rate movements so complicated. Meese and Rogoff (1983) argue that when using data from the 1970s, an RW model performs just as well as any conventional structural exchange rate model and some time series models. The greatest implication of the so-called "Meese–Rogoff puzzle" is that if what Meese and Rogoff (1983) propose were true, there would be no meaning in trying to forecast exchange rates at all. Since then, various studies have tried to explain the reason behind the puzzle (Rossi 2005; Clark & West 2007). For instance, Clark and West (2006, 2007) claim that the apparent poor performance of structural models against the RW is due to the wrong assumption that the MSPEs of nesting and nested models are expected to be the same under the null hypothesis.

Many have succeeded in using structural exchange rate models that perform better than the RW at certain levels (Mark 1995; Engel et al. 2007). Matsuki and Chang (2016) find that structural models, especially the Taylor rule, outperform the RW on short horizons with the USD/JPY rate. Time series models have been applied to exchange rate forecasting as well, such as by Mahmodpour et al. (2016). Hashimoto (2011) finds evidence of autoregressive moving average (ARMA) and vector autoregression (VAR) models outperforming the RW on short horizons with the Japanese yen. Given the moderate success of structural and time series models in the literature, in this study we use both types of models as benchmarks.

There is a growing literature on the use of Google Trends in the field of economics and finance. Onorante and Koop (2016) use search frequencies to improve the nowcasts of certain macroeconomic variables. Da et al. (2014) create a FEARS index with Google Trends to capture investor pessimism in the United States, finding that it is useful to explain price drops and volatility increases in the stock market. Chojnowski and Dybka (2017) succeed in picking up market sentiment in Poland with search frequencies and show that a VAR model including sentiment performs better than simple structural or RW models. Lastly, Bulut (2018) uses the search volumes of words related to macroeconomics to directly capture the market's perception of macroeconomic variables, producing forecasts that beat the RW for several currency pairs.

Complementing the above literature, we use Google Trends data to construct an SI similar to that of the FEARS index to capture market sentiment in Japan and the United States and apply it to exchange rate forecasting. There are several benefits to using Google as our source of information (rather than, for example, Yahoo! or Twitter). First, the data are readily accessible by anyone through the website. In addition, of all the search engines, Google has a dominant presence in both the countries[1] chosen for this study. Finally, when dealing with exchange rates, it is important to capture the collective sentiment of the population as a whole (as opposed to just investors on Twitter, for example), as exchange rates are known to be affected greatly by macroeconomic factors.

This study is in line with prior studies (Da et al. 2014; Bulut 2017), but extends the literature by exploring the possibility of using Google Trends data as a complementary tool to aid the long-lived difficult task of predicting exchange rates. In particular, we improve upon existing research in two main ways. First, we create an SI for Japan using words from a Japanese dictionary to prevent erroneous translations and capture market sentiment in the country as accurately as possible. Second, we show that the SI we propose has the potential to improve the exchange rate predictions of one-month-ahead USD/JPY returns.

3. Data and Methodology

3.1 Construction of the SI

In this study, we create an SI similar to the FEARS index of Da et al. (2014) as a proxy of market sentiment in the United States as well as a separate index to capture market sentiment in Japan. Below, we refer to the SI for the United States as the U.S. SI or SI_t^{US} and the SI for Japan as the Japanese SI or SI_t^{Japan}.

3.1.1 U.S. SI

To construct the U.S. SI, we follow the recent finance literature using text analysis

[1] According to StatCounter (https://statcounter.com/), Google has a market share of 86.9% in the United States and 74.4% in Japan (as of December 2018).

(Tetlock 2007; Da et al. 2014) and use the Harvard IV-4 Dictionary and Lasswell Value Dictionary. These dictionaries categorize English words according to their meaning or connotation, such as "active," "passive," "strong," and "weak." In this study, we are concerned with words related to the financial market, or "economic" words; hence, we extract the words with the "economic" tag. Moreover, since we are trying to capture market sentiment, we further filter the word list by choosing words that have a "positive" or "negative" tag in addition to the "economic" tag, resulting in a tentative list of 150 words. Da et al. (2014) include the top 10 searches for each of these words in their data; however, we decide not to do so since it results in numerous similar, overlapping keywords.[2]

For each of the words in the tentative list, we obtain monthly Google search volumes for January 2004 to August 2018, setting the geographic region to the United States. We download all the data at once in September 2018, and do not change or add any data to it.[3] Then, to prevent words with insufficient data from entering the index, we eliminate from this list any words that have an SVI of zero at any point in the sample period. This leaves us with a final list of 140 words, such as "inflation," "gold," "cheap," and "jobless."

Now, when we look at the SVI for each of these words, we can clearly see the presence of yearly seasonality. Figure 1 plots the monthly SVIs of the word "inflation;" there are relative spikes in the SVI around April each year and drops in the SVI around August. To address this problem of seasonality, we use the X-13 ARIMA-SEATS seasonal adjustment method[4] for each of the 140 words on the final list.

(Figure 1 here)

We define the monthly change in the search volume of term *i* as

$$\Delta SVI_{i,t} = \ln(SVI_{i,t}) - \ln(SVI_{i,t-1}). \tag{1}$$

Now, we find the correlations of $\Delta SVI_{i,t}$ with the USD/JPY rate and decide which words

---

[2] Google provides a list of related search terms. Although including these words does contribute to a thorough selection of words, many similar words end up in the final list, which we find unsuitable for this study. For example, in Da et al. (2014), of the top 10 terms strongly correlated with stock returns, four are related to gold: "gold prices," "gold price," "gold," and "price of gold."

[3] Google calculates the SVI from a random subset of data, leading to a slight difference in data downloaded at different times. Da et al. (2011) show that the SVI downloaded at different times has a correlation of 97% or higher. Although the effect of this sampling error is limited considering the aim of this study, we try to keep this error as small as possible by downloading all the necessary SVI data at once.

[4] This is the official method used by the U.S. Census Bureau; we choose this to match the seasonal adjustment methods of other macroeconomic data.

to include in the SI. Since our objective is to produce one-month-ahead forecasts, we calculate the correlations of $\Delta SVI_{i,t}$ and $(s_{t+1} - s_t)$, where s_t is the log of the monthly average USD/JPY spot rate. Panel A of Table 1 shows the 15 most positively and negatively correlated words during the initial training period (January 2004 to February 2011). In general, words that are positively correlated seem to have a positive image of the U.S. economy (e.g.,. luxury, success, gift), while negatively correlated words have a negative image (e.g., blackmail, jobless, shortage). Not all words seem to make sense in terms of explaining exchange rates. Nevertheless, for the purposes of this study, we select keywords as quantitatively and objectively as possible, and thus we do not manually add or remove any words.

(Table 1 here)

Tetlock (2007) and Da et al. (2014) find that in the English language, negative terms are most useful for identifying stock market sentiment. In the case of monthly exchange rates (USD/JPY), however, we find that this does not apply. Table 1 shows that both positively and negatively correlated words have a minimum absolute t-stat of 1.62, with positive words having slightly stronger t-statistics, most likely because of the complex nature of foreign exchange rates, which are affected by a wide range of factors in multiple countries.

Following the above results, we decide to include an equal number of positive and negative terms (15 of each). Thirty is considered to be sufficient to diversify idiosyncratic noise (Da et al. 2014); however, we also create indices with 20 and 10 words as a means of comparison.[5] We construct the index in the following manner.

A) For the initial training period (January 2004 to February 2011), we determine the 30 most correlated words (15 positive and 15 negative) with the USD/JPY rate, as shown in Table 1.

B) We define the U.S. SI, or SI_t^{US}, as

[5] Reducing the number of keywords risks introducing idiosyncratic noise to the index, but increasing the number risks introducing words with very weak correlations. Further, considering that many words beyond the cutoff of 30 are seemingly irrelevant to economics, we decide to test indices with fewer words.

$$SI_t^{US} = \sum_{i=1}^{15} \Delta SVI_{i,t} - \sum_{j=1}^{15} \Delta SVI_{j,t} \,. \tag{2}$$

$\Delta SVI_{i,t}$ denotes the log difference SVI for the term ranked *i* in the top 15 most positively correlated words and $\Delta SVI_{j,t}$ represents the same for the negatively correlated words. By reversing the sign of $\Delta SVI_{j,t}$, SI_t^{US} represents an SI that has a positive correlation with the USD/JPY rate.

C) For the training period and initial six months (March 2011 to August 2011), SI_t^{US} is defined by Equation (2) and the terms in Table 1.

D) For the SI_t^{US} of the next six months (September 2011 to February 2012), we update Table 1 by recalculating the correlations, expanding the period[6] to the most recent month (January 2004 to August 2011).

E) We continue to expand the window and update the word list every six months[7] until we have SI_t^{US} for the entire sample period (until August 2018).

3.1.2 Japanese SI

To create the Japanese SI, the most obvious choice would be to use the word list obtained in Section 3.1.1 and translate it into Japanese. Indeed, using the same word pool would allow for a direct comparison between the two countries. However, such a direct translation overlooks an important problem. English and Japanese are two very different languages with different roots; one word may have a completely different connotation in the other language, or may not exist at all. Bulut (2018), for example, creates an initial keyword list in English and uses Google Translate to translate the list into other languages, including Japanese. When we look at the Japanese translations, however, there are obvious mistranslations such as “spend” and “sugosu,” where the latter is used to describe the act of “spending time” as opposed to the intended meaning of “spending money.” For this reason, creating the Japanese SI from a separate Japanese dictionary is crucial to accurately capture

[6] Since the data are monthly and the number of data points is limited, we decide to expand the calculation window.

[7] The best choice may be to update the list every time we create a forecast. However, significant changes in the composition of the index could not be observed every month. Hence, we choose an update period of six months.

market sentiment in Japan.[8]

We use a dictionary provided by the NINJAL[9] Center of Corpus Development, which consists of a thorough list of Japanese words categorized based on their meanings. This dictionary provides numerous main categories such as “nature,” “life,” and “language,” with various subcategories for each of those. For our main category, we use “economy” (keizai), and “economy/balance of payments” (keizai/shushi) for our subcategory. Because this dictionary classifies “positive” and “negative” connotations like the Harvard IV-4 and Lasswell Value Dictionaries, we choose the subcategory that contains the most general economic terms (461 words). Setting the geographical region to Japan, we obtain the SVIs for each term on this tentative list for January 2004 to August 2018, removing words with insufficient SVI data (any word with an SVI of zero at any point). This leaves us with a final list of 186 words. Panel B of Table 1 shows the 15 most positively or negatively correlated words for January 2004 to February 2011.

We construct the Japanese SI in the same way as the U.S. one, defining it as

$$SI_t^{Japan} = \sum_{i=1}^{15} \Delta SVI_{i,t} - \sum_{j=1}^{15} \Delta SVI_{j,t}\,. \tag{3}$$

3.2 Structural exchange rate models

In this study, we use three structural exchange rate models following Bulut (2018) and Mark (1995) to act as base models; we add the SIs to them to examine whether they improve the forecast accuracy. The three models we use are the PPP model, IRP model, and monetary model. In theory, the PPP model is meant for use over long-term horizons, and thus it may seem unsuitable to apply this model to one-month-ahead forecasts. However, since Bulut (2018) uses these three models to obtain monthly forecasts and for comparison purposes, we decide to use them.

We use simple linear structural exchange rate models as in Mark (1995), which take the following form:

[8] One may argue that this undermines objectivity to a certain extent; however, we decide to emphasize preventing erroneous or subjective translations.
[9] National Institute for Japanese Language and Linguistics, dictionary available at: https://pj.ninjal.ac.jp/corpus_center/goihyo.html.

$$s_t = c + f_t. \tag{4}$$

s_t is the natural logarithm of the average monthly spot exchange rate, setting Japan as the home country (equivalent to the USD/JPY rate). Therefore, an increase in the value of s_t would signify the appreciation of the U.S. dollar and depreciation of the Japanese yen. f_t represents the fundamental value of the log exchange rate and c is a constant term representing the deviation of the actual log exchange rate s_t from its fundamental value f_t. Although such deviations can be persistent, the log exchange rate is expected to return to its fundamental value in the long run.

Next, we adopt the following regression equation used by Mark (1995) to forecast the exchange rates:

$$y_t = s_t - s_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_t. \tag{5}$$

Here, y_t represents the monthly log exchange rate returns $(s_t - s_{t-1})$ and x_{t-1} is the deviation of the log exchange rate from its fundamental value, $f_{t-1} - s_{t-1}$. As nominal exchange rates are known to be nonstationary unit root variables, differencing the series eliminates the need to worry about picking up any spurious relationships.

Once we specify Equation (5) for the three structural models, we add the variables SI_t^{Japan} and SI_t^{US} to improve the forecasts. Since we build these variables to have a positive correlation with log exchange returns, we simply include them in Equation (5) as additional linear regressors:

$$y_t = \alpha + \beta x_{t-1} + \gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US} + \varepsilon_t. \tag{6}$$

3.2.1 PPP model

The PPP model states that relative price levels determine the fundamental value of log exchange rates, or $f_t = p_t - p_t^*$, where p_t represents the natural log of price levels at home (Japan) and p_t^* represents the same value in the foreign country (United States). This implies that if inflation rates in the home country are relatively high, the purchasing power of the home currency will decrease, leading to the depreciation of that currency. The specifications for the PPP model from Equations (5) and (6) become the following:

$$y_t = \alpha + \beta(p_{t-1} - p_{t-1}^* - s_{t-1}) + \varepsilon_t, \tag{7}$$

$$y_t = \alpha + \beta(p_{t-1} - p_{t-1}^* - s_{t-1}) + \gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US} + \varepsilon_t. \tag{8}$$

3.2.2 IRP model

The IRP model states that returns on domestic currency assets are equal to the exchange rate-adjusted returns of foreign assets. In other words, if an investor borrows money in a country with a lower interest rate, exchanges it, and invests it in a foreign country with a higher interest rate, the currency in the foreign country will eventually depreciate because of the excess return gained from the difference in interest rates.[10] The fundamental value is defined by $f_t = i_t - i_t^* + s_t$, where i_t and i_t^* are the interest rates in the home and foreign countries, respectively. The regression models are expressed by the following:

$$y_t = \alpha + \beta(i_{t-1} - i_{t-1}^*) + \varepsilon_t, \tag{9}$$

$$y_t = \alpha + \beta(i_{t-1} - i_{t-1}^*) + \gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US} + \varepsilon_t. \tag{10}$$

3.2.3 Monetary model

This model defines exchange rates using the relative supply and demand of money at home and abroad. We use the same assumptions stated in Bulut (2018), where we assume that uncovered IRP and PPP hold and that both countries have the same income and interest rate elasticity of demand. The fundamental value in this case is defined by $f_t = m_t - m_t^* - \lambda(g_t - g_t^*)$, where m_t and m_t^* are the natural logs of money supply at home and abroad, g_t and g_t^* are the natural logs of real income, and λ is the income elasticity of money demand. As in Mark (1995) and Bulut (2018), we assume that $\lambda = 1$. The model specifications become the following:

$$y_t = \alpha + \beta(m_{t-1} - m_{t-1}^* - \lambda(g_{t-1} - g_{t-1}^*) - s_{t-1}) + \varepsilon_t, \tag{11}$$

$$y_t = \alpha + \beta(m_{t-1} - m_{t-1}^* - \lambda(g_{t-1} - g_{t-1}^*) - s_{t-1}) + \gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US} + \varepsilon_t. \tag{12}$$

[10] If the foreign currency stays the same, there would be infinite arbitrage opportunities, which the IRP model rejects.

3.3 Autoregressive models

In addition to the structural exchange rate models in Section 3.2, we use simple time series models as benchmarks and examine whether our SIs improve predictability. The basic concept behind time series models is that a certain variable is sometimes correlated with past values of itself, thereby allowing us to produce forecasts of that variable using its past values. Several studies including Hashimoto (2011) and Mahmodpour et al. (2016) have shown that time series models such as ARMA, ARIMA, and GARCH are effective to some extent at predicting future exchange rates. In this study, given our results from using the Durbin–Watson test and serial correlation Lagrange multiplier (LM) test,[11] we estimate a simple linear regression model with lagged dependent variables using ordinary least squares:

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t, \tag{13}$$

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t. \tag{14}$$

We use the above two specifications as the benchmark time series models, defining them as AR(1) and AR(2), respectively. To choose the number of lagged variables to include in the model, we consider the Bayesian and Akaike information criteria for lags up to four; we find that for this sample, the first- and second-order lags show better results. For each of these two models, we add up to two lagged variables of the SI, resulting in four new equations:

$$y_t = \alpha + \beta y_{t-1} + (\gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US}) + \varepsilon_t, \tag{15}$$

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + (\gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US}) + \varepsilon_t, \tag{16}$$

$$y_t = \alpha + \beta y_{t-1} + \left(\gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US}\right) + \left(\gamma_3 SI_{t-2}^{Japan} + \gamma_4 SI_{t-2}^{US}\right) + \varepsilon_t, \tag{17}$$

$$y_t = \alpha + \beta_1 y_{t-1} + (\gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US}) + \beta_2 y_{t-2} + \left(\gamma_3 SI_{t-2}^{Japan} + \gamma_4 SI_{t-2}^{US}\right) + \varepsilon_t\ . \tag{18}$$

3.4 Forecasting method

Our entire data set consists of monthly data for January 2004 to August 2018. We treat the first half of this period, January 2004 to February 2011, as the training sample and

[11] In this study, the null hypothesis of "no serial correlation" is not rejected at any level, and therefore we use standard ordinary least squares to estimate the equation.

produce 90 out-of-sample forecasts for March 2011 to August 2018. We use a recursive approach, expanding the window each month and re-estimating the parameters each time. Using Equation (13) as an example, to produce the first forecast for March 2011, we first estimate the parameters $\hat{\alpha}$ and $\hat{\beta}$ using the training sample (January 2004 to February 2011). Then, we use those parameters to produce the forecast $\hat{y}_{t+1}$ (equivalent to $\hat{s}_{t+1} - \hat{s}_t$) using the obtained parameters: $\hat{y}_{t+1} = \hat{\alpha} + \hat{\beta} y_t + \varepsilon_t$. To create the next forecast $\hat{y}_{t+1}$ for April 2011, we re-estimate the parameters $\hat{\alpha}$ and $\hat{\beta}$ with an expanded sample window of January 2004 to March 2011 and obtain $\hat{y}_{t+2} = \hat{\alpha} + \hat{\beta} y_{t+1} + \varepsilon_t$.[12] This method is applied to all the models mentioned above, specifically Equations (7)–(18).

3.5 Forecast evaluation

In this section, we discuss the quantitative methods we use to evaluate our forecasts. Specifically, we use the MSPE, CW test, and DOC test following Bulut (2018). Additionally, we compare the accuracy of our models with a benchmark RW model without drift.

3.5.1 RW model

The RW model without drift is a nonstationary process defined by the following equation:

$$Y(t+1) = Y(t) + \alpha, \tag{19}$$

where α represents white noise.[13] In this form, the best one-step-ahead forecast of $Y(t)$ is the value of $Y(t)$ itself. When we difference the sequence and make it stationary, the equation becomes the following:

$$Y(t+1) - Y(t) = d(t) = \alpha. \tag{20}$$

This implies that $d(t+1) = d(t) = \alpha$, signifying that the best estimate of future returns using the RW model is zero, since α has a mean of zero.

---

[12] As we are producing static forecasts, we use the actual value y_{t+1} instead of the forecasted value $\hat{y}_{t+1}$ to calculate $\hat{y}_{t+2}$.

[13] White noise is a discrete signal whose mean is 0, with finite variance and no serial correlation.

3.5.2 MSPE

The simplest means of evaluating prediction results is by calculating the MSPE. We calculate the prediction error pe_i of each forecast by $pe_i = y_{t+1} - \hat{y}_{t+1}$ and define the squared prediction error as $spe_i = (y_{t+1} - \hat{y}_{t+1})^2 = pe_i^2$. Finally, the MSPE can be described as

$$MSPE = \frac{1}{k}\sum_{i=1}^{k} spe_i\,. \tag{21}$$

3.5.3 CW test

Clark and West (2006, 2007) propose a method of testing the equal predictive accuracy of two nested models. A model is nested in another model when its parameters are a subset of the parameters in the second model. For example, let us look at the following models:

$$y_t = \varepsilon_t, \tag{22}$$

$$y_t = \alpha + \beta x_{t-1} + \varepsilon_t, \tag{23}$$

$$y_t = \alpha + \beta x_{t-1} + \gamma_1 SI_{t-1}^{Japan} + \gamma_2 SI_{t-1}^{US} + \varepsilon_t. \tag{24}$$

Equation (22) represents the differenced series of an RW model, where ε_t is analogous to white noise. Equations (23) and (24) represent a structural exchange rate model and a structural exchange rate model with additional SI parameters, respectively. Here, Equation (22) can be interpreted as a "nested" model within Equation (23), while Equation (23) is also nested within Equation (24).

Clark and West (2007) suggest that when comparing the predictive accuracy of nested models such as Equations (22) and (23), the MSPE of the larger model is expected to have a greater MSPE under the null that the parsimonious (original) model is true. This is because if the parsimonious model is true, the larger model introduces noise into its forecasts by using irrelevant variables (population values of these parameters should be zero). The implication is that when comparing two nested models, the MSPEs should not be compared directly, and rather an adjusted version of the MSPE should be used.

The CW test is an adjusted version of the Diebold and Mariano (1995) test for equal predictive accuracy, which conducts a one-tailed significance test for the null hypothesis that the MSPEs of two models are the same. We define Model p as the parsimonious model and Model i as the larger model that nests Model p. Let $spe_p = (y_{t+1} - \hat{y}_{t+1}^p)^2$ and $spe_i = (y_{t+1} - \hat{y}_{t+1}^i)^2$, where $\hat{y}_{t+1}^p$ and $\hat{y}_{t+1}^i$ are the forecasts produced by Models p and i, respectively. Diebold and Mariano (1995) test the null hypothesis that $E[spe_p - spe_i] = 0$, whereas Clark and West (2007) adjust spe_i to account for the noise produced by defining an adjustment term: $adj = (\hat{y}_{t+1}^p - \hat{y}_{t+1}^i)^2$. The new null hypothesis to test becomes

$$H_0: E\left[\hat{f}_t = spe_p - (spe_i - adj)\right] = 0. \tag{25}$$

The spe_i for the larger model is adjusted downward to account for noise. Clark and West (2007) show that f_t approximately follows a normal distribution with a mean of zero. To test this null hypothesis, we use the CW test statistic defined by the following, where P is the number of forecasts:

$$CW = \frac{E[\hat{f}_t]}{\sqrt{\frac{\sum_{i=1}^{P}\left(\hat{f}_t - E[\hat{f}_t]\right)^2}{P}}}. \tag{26}$$

To perform a one-tailed test, we reject the null when the CW statistic is greater than 1.282 (for 10% significance) or 1.645 (for 5% significance). We perform this test for all of our models against the RW model as well as for all of our models with SIs against their original structural or time series models. When Model p is the RW, the one-step-ahead forecast is zero, and thus $spe_p = (y_{t+1})^2$ and $adj = (\hat{y}_{t+1}^i)^2$.

3.5.4 DOC test

Apart from looking at the quantitative deviations of the forecast values from the actual values, we also examine how well our models predict the direction of exchange rate movements. Under the RW null, the best one-step-ahead forecast is the current exchange rate, or in other words no change at all. As the RW model does not predict the DOC, we represent this situation with a simple “coin toss” model following Diebold and Mariano (1995) and Bulut (2018); the RW model is thought to have an equal chance of moving in the same or opposite direction of the actual exchange rate. The value $\overline{DOC}$ is defined by the sample average of the following function, where $S(y_{t+1})$ is a sign function that takes a value of one for positive numbers and zero for negative numbers:

$$DOC = \left[S(y_{t+1})S\left(\hat{y}_{t+1}^{i}\right)\right] + \left[1 - S(y_{t+1})\right]\left[1 - S\left(\hat{y}_{t+1}^{i}\right)\right]. \tag{27}$$

Here, the value DOC takes a value of one when the forecast matches the sign of the actual exchange rate return and zero when it does not. The average, $\overline{DOC}$, represents the proportion of forecasts that correctly predicted the sign of the exchange rate movements. Our benchmark RW model (coin toss model) takes the value $\overline{DOC} = 0.5$, as the probability of correctly predicting the sign of returns is 0.5. To perform the DOC test, we use the DOC test statistic defined by the following:

$$DOC\ \ stat = \frac{\overline{DOC} - 0.5}{\sqrt{\frac{0.25}{N}}}. \tag{28}$$

Here, N is the number of forecasts and the $DOC\ \ stat$ follows a standard normal distribution.

3.6 Macroeconomic data

We obtain all of our official macroeconomic data from the Federal Reserve Bank of St. Louis, Statistics Bureau of Japan, the Bank of Japan, and the Ministry of Economy, Trade, and Industry of Japan. Exchange rates (USD/JPY) are obtained as monthly average values, where Japan is set as the home country; an increase in the rate implies a depreciation of the yen.

Price levels (p_t) are defined by the natural logarithm of the consumer price indices of each country. For interest rates (i_t), we use the one-month Treasury bill rate for the United States and the one-month JPY LIBOR interest rate for Japan. The natural log of M1 is used to represent the money supply (m_t) of each country and the Industrial Production Index is used as a proxy for total income (g_t).

4. Results

4.1 MSPE, CW, and DOC results for the structural models

Table 2 presents the results related to the structural models and results for the time series models. Panels A to C show the results for the PPP, IRP, and monetary models, respectively. We include the results for SIs with a total word count of 30, 20, and 10 as a means of comparison. In addition to the MSPEs of each model, we show the percentage improvement in the MSPE for each model compared with the RW and base structural model. The MSPE for the RW model is 5.27×10^{-4}, and this will serve as a benchmark for all the models. For the CW and DOC tests, we show the t-statistic for each test, where they approximately follow a standard normal distribution.

(Table 2 here)

Table 2 shows that none of the structural models beats the RW model in terms of the MSPE or DOC. This is true even when we use the CW test to account for the expected upward shift in the MSPE of the nesting models. Our results are not surprising given that structural exchange rate models are typically more adequate when dealing with much longer horizons (such as quarter-ahead or year-ahead forecasts), as Mark (1995) does. Our results are not consistent with Bulut (2018), who finds that the PPP and monetary models for Japan outperform the RW null when looking at the CW statistics. We presume this difference may come from technical differences in the method and period of forecasts.

Next, we examine the effects of the 30-word, 20-word, and 10-word SIs on the forecast accuracy. The three models show similar results; the MSPEs improve slightly (by about 0.9%) when using the 30-word index, but worsen for the 20-word and 10-word indices. As for the DOC, there are no statistically significant results; however, the 30-word index seems to perform slightly better than the other two. Furthermore, the DOC is negative for all three structural models (fewer than half of the forecasts predict the direction correctly), whereas the addition of the 30-word index improves this to a positive number for all the models. This superiority of the 30-word index is consistent with the fact that 30 is often considered to be the minimum number needed to diversify away idiosyncratic noise (Da et al. 2014). In other words, using too few words to compose an index introduces more noise from individual words than it does predictability as a whole.

4.2 Regression results for the structural models

Table 3 presents the regression results of the structural models used to produce the forecasts to obtain a different perspective on the performances of each model. Panels A to C show the results of the PPP, IRP, and monetary models, respectively. The regressions are conducted using the entire data set (January 2004 to August 2018).

(Table 3 here)

First, as expected from the relatively poor forecasting ability of the structural models, none of the price-related, interest rate-related, or money supply-related coefficients shows a significant t-statistic. In fact, even when looking at the other window sizes,[14] none of the estimated coefficients shows statistical significance. On the contrary, all the SIs in the three models show statistical significance at the 1% level, explaining their contribution to the improvement in forecasting ability; the adjusted R-squared values[15] and standard errors (SEs) also improve with their presence. Furthermore, in Section 4.1, when focusing on the MSPE as a measure of predictive accuracy, the 30-word SI shows the best results. The same is true for the regression results; the adjusted R-squared value is the highest for each structural model when the 30-word index is added as a predictor and the SE of the regression is the lowest for those models as well.[16]

4.3 MSPE, CW, and DOC results for the time series models

Table 4 presents the results of the models defined by Equations (15)–(18). Panels A to C show the MSPEs and t-statistics of the CW/DOC tests for the AR models and SI(30, -1) to SI(10, -1), respectively. Contrary to the relatively poor performance of the structural models, the base time series models AR(1) and AR(2) perform considerably better; they show a decrease in the MSPE of 6.09% and 6.99% compared with the RW, respectively. The CW statistics are both significant at the 5% level, meaning that the null hypothesis that exchange rates cannot be explained is rejected at 5%. This result is consistent with the findings of past studies (Hashimoto 2011; Mahmodpour et al. 2016).

[14] The forecasts are produced using an expanding estimation window, where the coefficients are estimated each time a forecast is made. Therefore, 90 sets of coefficients exist for each model; the structural models do not have any significant coefficients in any of the 90 regressions.
[15] The adjusted R-squared value takes into account the fact that when new variables are introduced into a model, the R-squared value automatically increases. The adjusted R-squared increases only if the increase in R-squared due to the addition of a new variable is more than that expected by chance.
[16] The coefficients of the SI are statistically significant for all the other window sizes as well, with at least 5% significance.

(Table 4 here)

We obtain good results for the SIs as well. Looking at the MSPEs, the 30-word SI performs best, improving the predictive power of AR(1) in both cases and AR(2) in one. As for the 20- and 10-word indices, only one model of the eight shows an improvement relative to its base model. Together with the results in Section 4.1, we can thus clearly see a difference in performance, most likely due to the noise present in indices with too few constituent words. The CW statistics show significance at the 5% level or higher for all the models compared with the RW, but do not show significance compared with AR(1) and AR(2).

Looking at the DOC statistics, the 20- and 10-word indices do not show any statistical significance, whereas the 30-word index shows a significant improvement in three of the four models, once again proving its superiority. Here, the DOC is not statistically significant for the base AR(1) and AR(2) models; however, when the SI is added into the model, it becomes significant. This clearly shows that to some extent, the index has the power to help predict whether the exchange rate will rise or fall in the next month.

4.4 Regression results for the time series models

Table 5 shows the regression results for the time series models. Panels A and B present the results for the AR(1) and AR(2) models, respectively. In general, we verify that the time series models are a useful way of describing exchange rate movements. More specifically, the one-month lag of exchange rate returns is statistically significant in all the models, while the two-month lag does not show statistical significance anywhere. However, when comparing the base AR(1) and AR(2) models, we see that the addition of the two-month lag increases the adjusted R-squared. Further, although the coefficient does not reach statistical significance, the t-value is fairly high (p=0.11). We see that once again, as with the structural models, the SIs improve the AR(1) and AR(2) models in terms of the adjusted R-squared and SE of the regression. In addition, models using the 30-word index as additional predictors show the best adjusted R-squared and SE of the regression.

(Table 5 here)

When looking at the individual variables, the U.S. SI is statistically significant at the

1% level for all the models, while the Japanese SI shows slightly less significance (5% or 10%) in some models. We do not see this as a problem since the Japanese SI does not lose statistical significance anywhere. However, we suspect that this small difference may come from the fact that we use different dictionaries for each language,[17] or perhaps due to a fundamental difference in the effectiveness of picking up investor sentiment via Google Trends in the two countries.

For a different perspective on how the AR(1)+SI(30, -1) and AR(2)+SI(30, -1) models behave throughout the forecasting period, we graph the p-values of all the dependent variables, excluding the constant. Figures 2 and 3 show that the base autoregressive models explain the data well; the one-month lag of returns is statistically significant for most of the period, and although the two-month lag is only significant in several places, it falls mostly in the p=0.1 to p=0.2 range when it is not.

(Figures 2 and 3 here)

Figure 4 shows that all three dependent variables have a p-value of less than 0.01 for the entire forecasting period, indicating a good and robust fit of the AR(1)+SI(30, -1) model. Figure 5 reports that the explanatory power of the two-month lag of returns decreases compared with AR(2) with the introduction of the SIs. Compared with Figure 4, the one-month lag of returns, Japanese SI, and U.S. SI behave in the same way. This indicates that the addition of a second lag of returns does not improve the model by a significant amount and that most of the predictive power is derived from the AR(1)+SI(30, -1) model.

(Figures 4 and 5 here)

Together with the findings with the structural model, we can thus conclude that the SI has the potential to improve the predictive ability of exchange rate models. The effects are more prevalent in the time series models, as the SI improves those already significantly performing better than the RW. However, this is limited to the 30-word SI; the 20- and 10-word indices clearly perform worse.

[17] When creating the Japanese index, we limit our original word list to keywords in the "economy/balance of payments" section within the general "economy" section rather than using all the words with the "economy" tag as we do with the English index. This may have led to a slightly biased Japanese index compared with the U.S. index, resulting in lower p-values in some models.

5. Robustness Checks

In the above section, we find that when comparing the performance of the 30-word, 20-word, and 10-word indices, the 30-word index shows the best results. We conduct additional analyses by creating three additional types of indices with 25, 35, and 40 words[18] to examine whether we can gain more insight into the behavior of the SI. We find that for these structural models, there is some evidence of forecast improvement when using the 25-word index (0.03–0.16%) and no sign of improvement with the other indices.

For the time series models, the 25-word index produces the best results as well. Although the index makes the MSPE of AR(1) increase slightly, the addition of the index raises the DOC to a statistically significant level. Further, SI(25, -1) improves AR(2) by 2.01%, leading to an 8.87% improvement over the RW model. The 35- and 40-word indices generally decrease the accuracy of AR(1) and AR(2) by several percent, although all the models perform better than the RW because of the good initial performance of both models.

(Table 6 here)

Table 6 summarizes our findings on the difference in the performance of the SIs depending on the number of keywords included in the index. We determine that an index helped "improve" a model when its addition either decreased the MSPE of the base model or increased the DOC statistic to a statistically significant level. Although this is a relatively rudimentary way of comparing the indices, it is clear that the 25- and 30-word indices outperform the others. As for the 30-word index, we find that it helps improve the forecast accuracy to a certain degree in *all* the models that we test in this study. From this, we conclude that too few words composing the index leads to noise due to a lack of diversity, as Da et al. (2014) point out. However, as one may easily infer, when we add more words to the index, the words' correlations with the market weaken. In our case, where the sample keyword list is a few hundred words,[19] the advantage of diversification seems to overcome the disadvantage of adding weakly correlated words at around 25–30 words.

---

[18] For the 25-word index, we use the 13 most positively correlated and 12 most negatively correlated words, while for the 35-word index, we use the 18 most positively correlated and 17 most negatively correlated words to construct the index.

[19] If the original keyword list is much larger (e.g., 3000 words), the threshold word count might be higher as more words would be highly correlated with the market.

6. Concluding Remarks

In this study, we explore the usage of Google Trends as a way of picking up market sentiment in Japan and the United States to help predict the exchange rate returns between these two countries. To do so, we construct two SIs for Japan and the United States using dictionaries in their native language and include them as additional regressors in various models to examine whether they improve the predictive accuracy of the original models.

We use three major structural models following Bulut (2018) and two time series models as the base models as well as consider the RW model as a general benchmark. We find that when adding the SI to these models, the MSPE and DOC forecasts improve in many cases, especially when using the 30-word SI. For example, the AR(1) and AR(2) models show a 6.09% and 6.99% decrease in the MSPE compared with the RW, and when the SI is introduced, these percentages rise to 6.38% and 8.65%, respectively. Furthermore, in the case of AR(1)+SI(30, -1), the ability of the model to predict directional changes in exchange rate returns improves by a statistically significant amount. The regression results also reveal the effect of the index, improving the adjusted R-squared values and SEs of the regression in all cases.

We also conduct additional analyses using different numbers of total words included in the SI. We test six types of indices, finding that the 25- and 30-word indices produce especially good results overall. As for the 30-word index, it improves all seven of the models that we use in this study to a certain extent. Da et al. (2014) mention that the typical word count needed to diversify away idiosyncratic noise is 30, and we see this clearly from our results. Building on to this, we can infer from our findings that simply adding more words to the index does not improve effectiveness; at some point, the disadvantage of weakly correlated words entering the index exceeds the advantage of diversification. In our research, this optimal point is 30.

We would like to point out several possible improvements for future consideration. First, it may be useful to consider other structural models as reference models, as the models used in this investigation failed to beat the RW. The SIs improve these models to a certain extent, but the argument is somewhat weak considering that the original model has the same accuracy as the RW. We chose to use these models as a loose comparison to Bulut (2018); if other structural models describe the USD/JPY rate better,[20] it may be worth trying those.

Second, conducting this research with a different timespan and different window lengths may help us obtain more insight. In this study, we use the monthly exchange rate returns and monthly search volumes of keywords; hence, daily or weekly analyses could be performed to capture more detailed movements. Considering the possibilities of real-world applications and to take advantage of the readily available online data, analyses over shorter timespans are crucial.

Third, it may be useful to consider other ways of constructing the SI, especially the Japanese one.[21] The dictionary we use to construct the Japanese SI simply contains too many words under the category "economic." Therefore, we picked a subcategory that seemed to include the most general economy-related words. However, this process may have led to us omitting keywords that otherwise could have helped increase the predictive power of the SI. Using a different dictionary, categorizing words using original algorithms, and/or simply trying to use the entire pool of "economic"-related words may be ways to improve the index.

Lastly, we understand that the prediction of exchange rate movements has been a long-lasting difficult research field because of its complex nature. We do not think that data from Google Trends alone will solve this problem, but we hope that it will eventually serve as one of many powerful tools for helping predict exchange rates.

Acknowledgements

We would like to thank Editage (www.editage.jp) for English language editing. All remaining errors are our own.

---

[20] The main purpose of this study was to examine the effect of Google Trends data on the forecast accuracy of exchange rate models; we did not focus on searching for a perfect structural model that explains the USD/JPY rate.

[21] We say this since the Japanese SI shows slightly less predictive power than the U.S. SI.

References

Adachi, Y., Masuda, M., & Takeda, F. (2017). Google search intensity and its relationship to the returns and liquidity of Japanese startup stocks. *Pacific-Basin Finance Journal, 46*, 243-257.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management, 25*(3), 239-264.

Bulut, L. (2018). Google Trends and the forecasting performance of exchange rate models. *Journal of Forecasting, 37*(3), 303-315.

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*, 2-9.

Chojnowski, M., & Dybka, P. (2017). Is exchange rate moody? Forecasting exchange rate with Google Trends data. *Econometric Research in Finance*, *2*, 1-21.

Clark, T.E., & 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, T.E., & West, K.D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, *138*(1), 291-311.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461-1499.

Da, Z., Engelberg, J., & Gao, P. (2014). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, *28*(1), 1-32.

De Long, J B., Shleifer, A., Summers, L.H., & Waldmann, R.J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703-738.

Diebold, F.X., & Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253.

Engel, C., Mark, N., & West, K. (2007). Exchange rate models are not as bad as you think. *NBER Macroeconomic Annual*, *22*, 381-441. doi:10.3386/w13318

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012-1014.

Hashimoto, J. (2011). How well can the time series models forecast exchange rate changes? *Bulletin of Niigata Sangyo University*, *Faculty of Economics*, *39*, 1-15.

Mahmodpour, K., Badooei, Y.S., Mohseni, H., & Veismoradi, S. (2016). The comparative comparison of exchange rate models. *International Journal of Economics and Financial Issues*, *6*(2), 380-385.

Mark, N.C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability. *American Economic Review*, *85*(1), 201-218.

Matsuki, T., & Chang, M. (2016). Out-of-sample exchange rate forecasting and macroeconomic fundamentals: The case of Japan. *Australian Economic Papers*, *55*(4), 409-433.

Meese, R.A., & Rogoff, K. (1983). Empirical exchange rate models of the seventies. *Journal of International Economics*, *14*(1-2), 3-24.

Onorante, L., & Koop, G. (2016). Macroeconomic nowcasting using google probabilities. *Proceedings of the 1st International Conference on Advanced Research Methods and Analytics*.

Rossi, B. (2005). Testing long-horizon predictive ability with high persistence, and the Meese-Rogoff puzzle. *International Economic Review*, *46*(1), 61-92.

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.

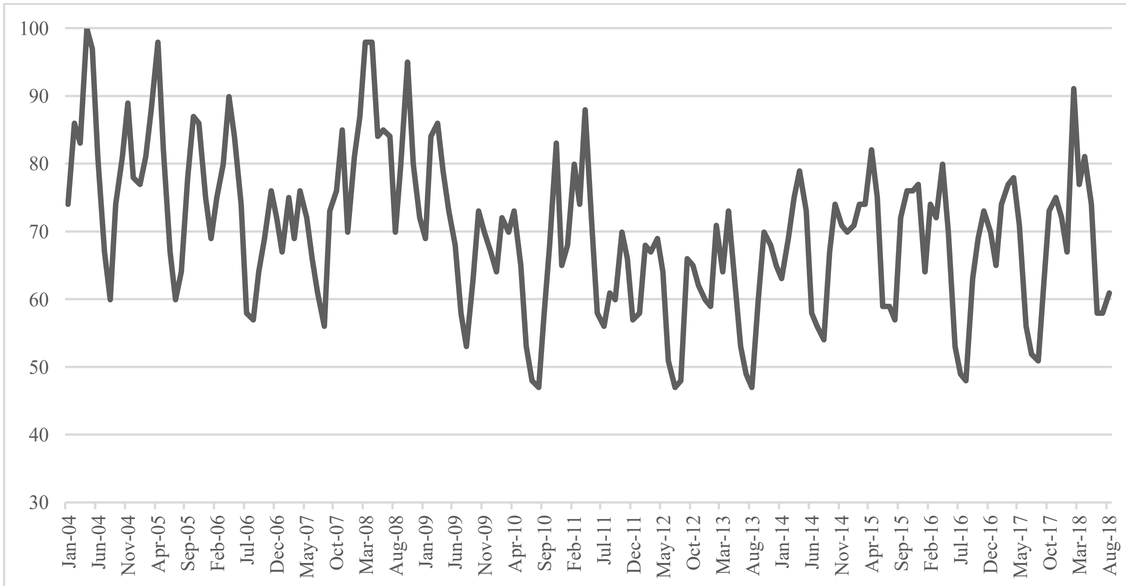

Figure 1: Monthly SVI data for "inflation"

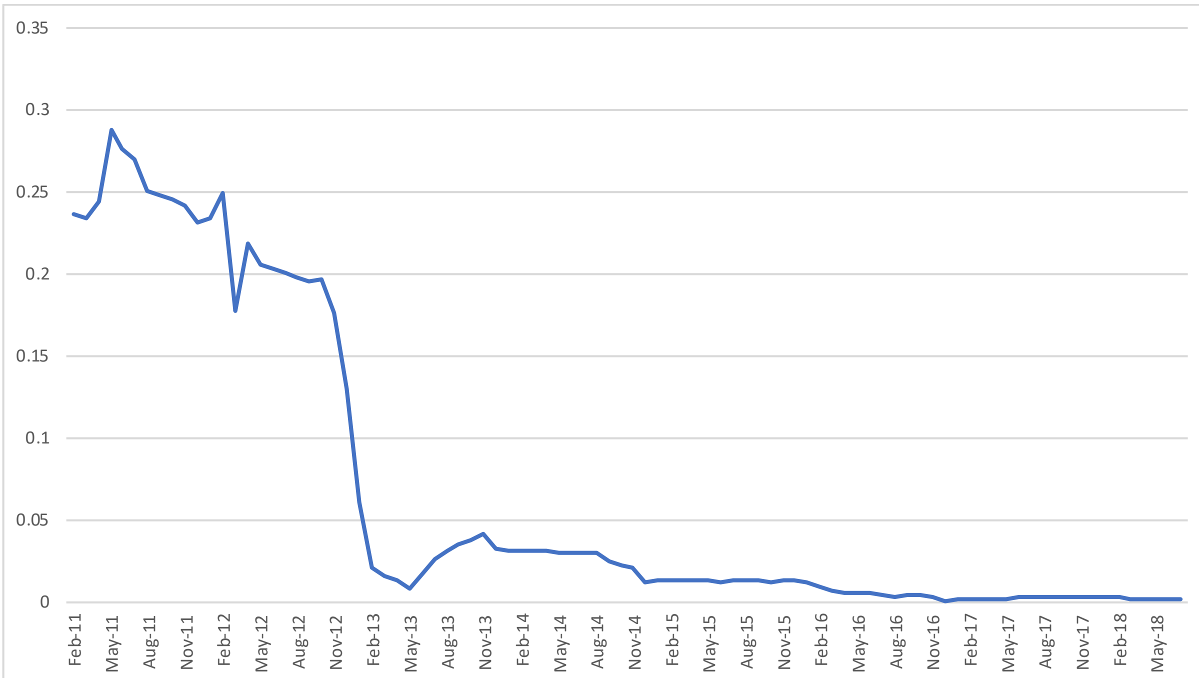

Figure 2: p-values for AR(1)

This graph plots the p-values for the one-month lag of exchange rate returns, DOLLAR_YEN(-1). The dates on the x-axis denote the end of the estimation window (e.g., Feb-11 refers to the regression from January 2004 to February 2011 used to forecast returns for March 2011).

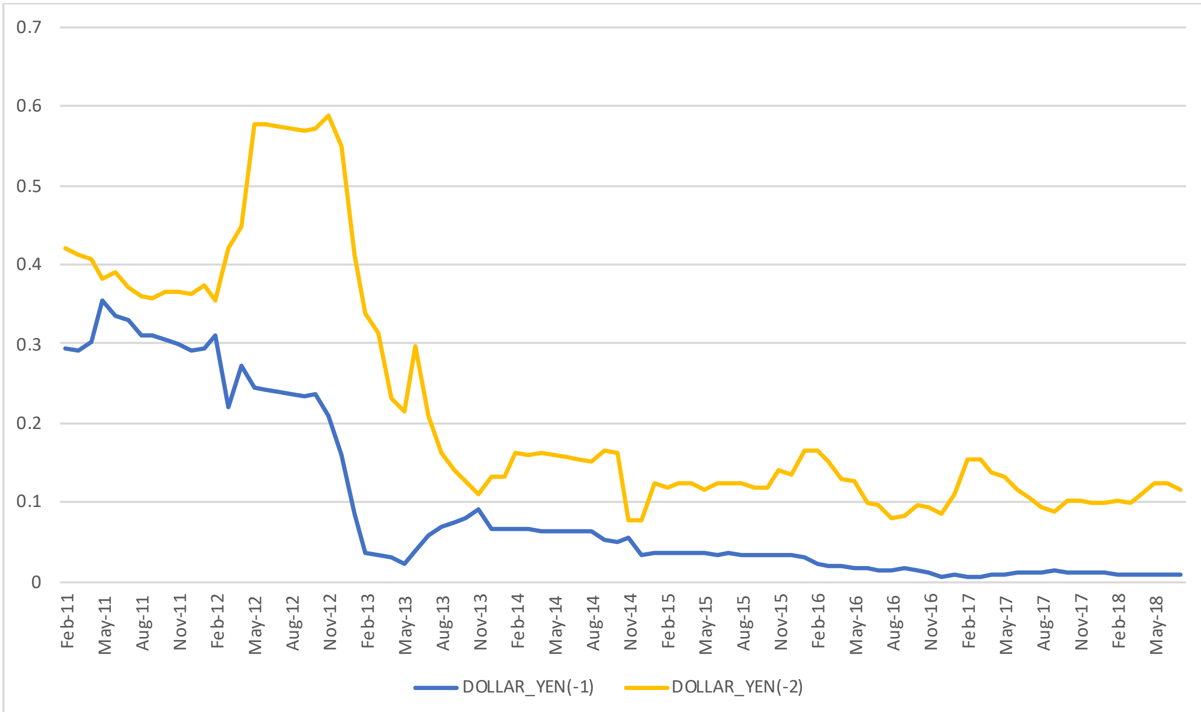

Figure 3: p-values for AR(2)

This graph plots the p-values for the one-month and two-month lags of exchange rate returns. The dates on the x-axis denote the end of the estimation window (e.g., Feb-11 refers to the regression from January 2004 to February 2011 used to forecast returns for March 2011).

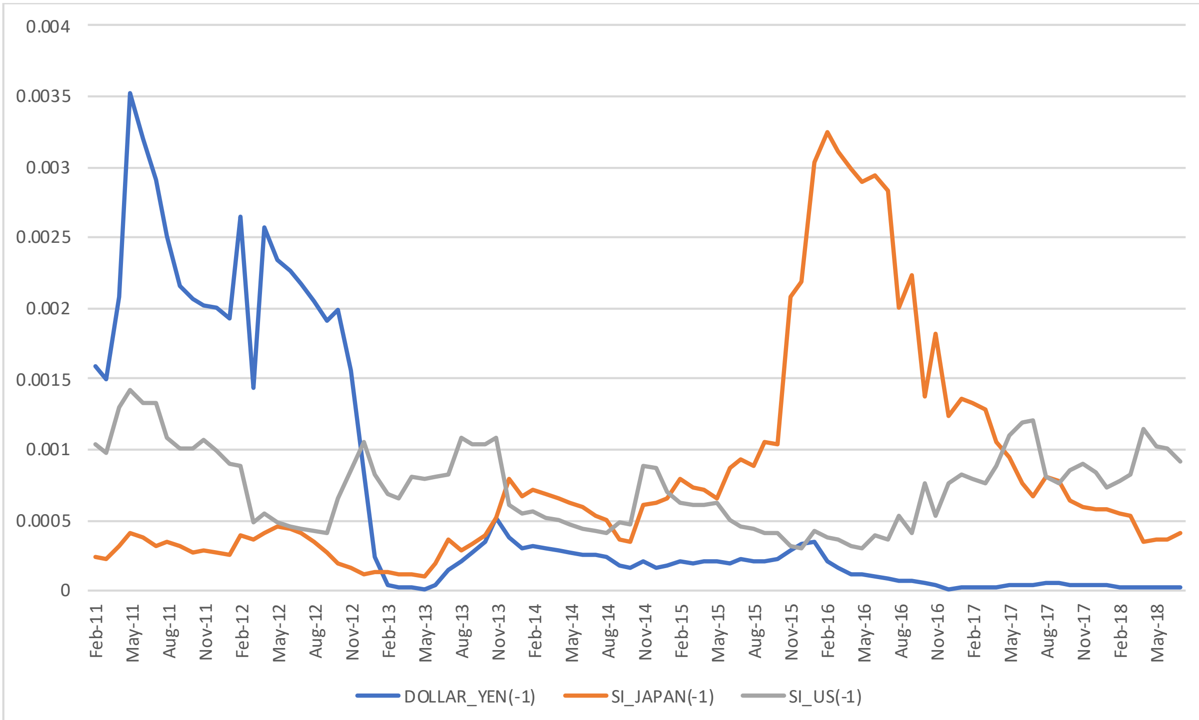

Figure 4: p-values for AR(1)+SI(30, -1)

This graph plots the p-values for the dependent variables in AR(1)+SI(30, -1). The dates on the x-axis denote the end of the window (e.g., Feb-11 refers to the regression from January 2004 to February 2011 used to forecast returns for March 2011).

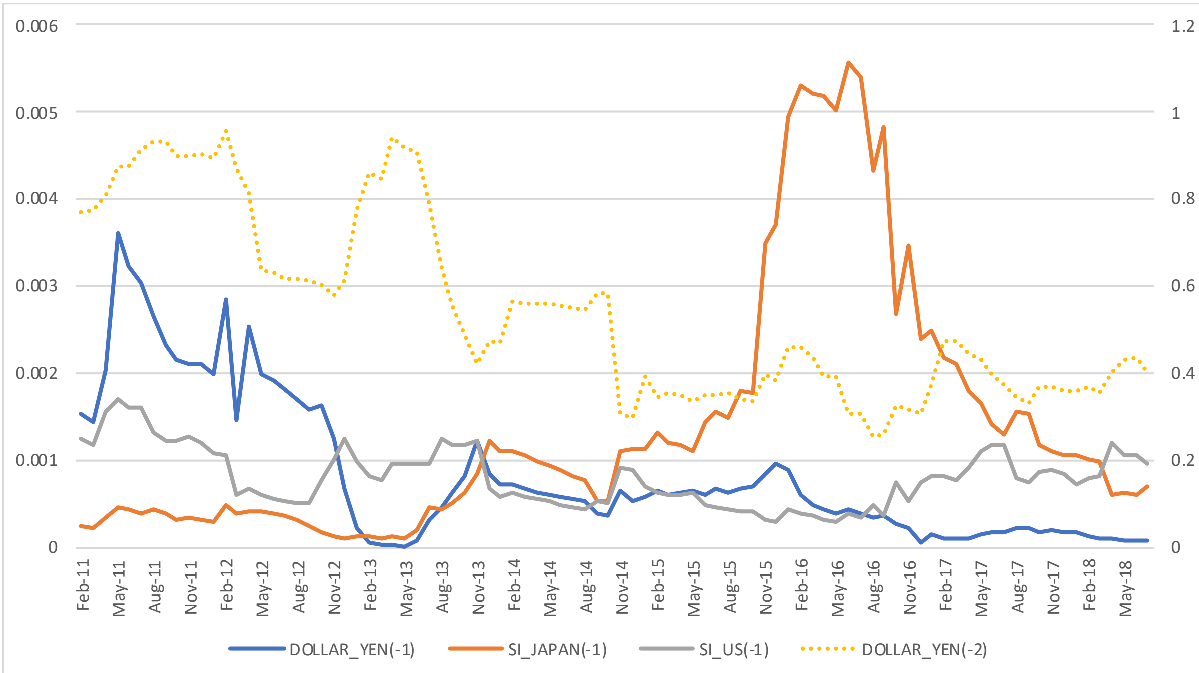

Figure 5: p-values for AR(1)+SI(30, -1)+SI(30, -2)

This graph plots the p-values for the dependent variables in AR(1)+SI(30, -1)+SI(30, -2). The dates on the x-axis denote the end of the window (e.g., Feb-11 refers to the regression from January 2004 to February 2011 used to forecast returns for March 2011).

Table 1: Terms included in the U.S./Japan SI for the initial training period

Panels A and B show the terms included in the U.S. and Japanese SIs, respectively for the initial training period. The two panels present the 15 most positively or negatively correlated terms with the USD/JPY rate for the initial training period: January 2004 to February 2011.

Panel A: Terms included in the U.S. SI for the initial training period

	Positive Correlation			Negative Correlation	
rank	word	t-statistic	rank	word	t-statistic
1	INFLATION	3.33	1	BLACKMAIL	-2.96
2	INTERVENTION	3.20	2	PRIVILEGED	-2.64
3	GOLD	3.10	3	GHETTO	-2.57
4	GAMBLE	2.84	4	EXTRAVAGANT	-2.42
5	LUXURY	2.58	5	NOBILITY	-2.41
6	RUIN	2.49	6	CHEAP	-2.29
7	ASSOCIATE	2.45	7	ENTREPRENEURIAL	-2.27
8	COOPERATIVE	2.26	8	JOBLESS	-2.16
9	POLLUTION	2.21	9	SHORTAGE	-2.04
10	SUCCESS	2.19	10	FELLOWSHIP	-1.98
11	GIFT	2.09	11	INEXPENSIVE	-1.92
12	AFFLUENT	2.08	12	VALUABLE	-1.92
13	BENEVOLENCE	2.07	13	BARGAIN	-1.73
14	PARTNERSHIP	2.07	14	COLONY	-1.68
15	POOR	1.98	15	LIQUIDATION	-1.62

Panel B: Terms included in the Japanese SI for the initial training period

	Positive Correlation			Negative Correlation	
rank	**word**	**t-statistic**	**rank**	**word**	**t-statistic**
1	出資	3.68	1	高価	-3.79
2	マネーゲーム	3.65	2	労災	-3.28
3	収支	3.39	3	口過ぎ	-3.18
4	倹約	3.14	4	消費する	-2.44
5	保険	2.92	5	収入	-2.25
6	インフレ	2.90	6	傷害保険	-2.21
7	デフレ	2.85	7	不要	-2.20
8	健保	2.82	8	後払い	-2.20
9	高め	2.81	9	未収	-2.19
10	結構	2.80	10	税収	-2.03
11	可処分所得	2.73	11	賭博	-1.99
12	特別会計	2.48	12	投機	-1.90
13	賭け	2.35	13	勝手向き	-1.83
14	決算	2.30	14	減資	-1.79
15	なくてもいい	2.22	15	高い	-1.73

Table 2: MSPE and CW/DOC results for the PPP, IRP, and monetary models

Panels A to C show the MSPEs and t-statistics of the CW/DOC tests for the PPP, IRP, and monetary models, respectively. SI(30, -1), SI(20, -1), and SI(10, -1) stand for the 30-word SI with a lag of 1, 20-word SI, and 10-word SI, respectively. The improvement in the MSPE is defined by the percentage decrease in the MSPE compared with the base model. As for the CW/DOC tests, none of the values showed a significant improvement.

Panel A: PPP model	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
MSPE (e-04)	5.484	5.434	5.703	5.694
Improvement (vs. RW)	-3.98%	-3.03%	-8.14%	-7.96%
Improvement (vs. PPP)	—	0.92%	-4.00%	-3.83%
CW and DOC tests				
CW-stat (vs RW)	-2.370	0.308	-0.370	-0.240
CW-stat (vs PPP)	—	1.035	0.274	0.378
DOC-stat	-0.843	1.054	-0.211	0.000

Panel B: IRP model	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
MSPE (e-04)	5.529	5.482	5.753	5.739
Improvement (vs. RW)	-4.83%	-3.95%	-9.08%	-8.82%
Improvement (vs. IRP)	—	0.84%	-4.06%	-3.81%
CW and DOC tests				
CW-stat (vs RW)	-1.203	0.357	-0.386	-0.231
CW-stat (vs IRP)	—	0.903	0.192	0.317
DOC-stat	-1.476	0.632	0.632	0.422

Panel C: Monetary model	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
MSPE (e-04)	5.510	5.463	5.734	5.725
Improvement (vs. RW)	-4.47%	-3.59%	-8.72%	-8.54%
Improvement (vs. Monetary)	—	0.85%	-4.06%	-3.89%
CW and DOC tests				
CW-stat (vs RW)	-2.424	0.220	-0.438	-0.300
CW-stat (vs Monetary)	—	1.003	0.255	0.362
DOC-stat	-0.843	0.211	-0.632	-0.211

Table 3: Regression results for the PPP, IRP, and monetary models

The regression results for the PPP, IRP, and monetary models are shown. SI(30, -1), SI(20, -1), and SI(10, -1) stand for the 30-word SI with a lag of 1, 20-word SI, and 10-word SI, respectively. The independent variable is the log difference in monthly USD/JPY returns.
Note: *, **, and *** denote the 10%, 5%, and 1% levels of significance, respectively.

Panel A: PPP model

	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Intercept	0.044	0.032	0.033	0.033
	(0.63)	(0.49)	(0.50)	(0.49)
$p_{t-1} - p_{t-1}^{*} - s_{t-1}$	0.019	0.013	0.014	0.014
	(0.62)	(0.49)	(0.50)	(0.49)
SI_{t-1}^{Japan}		0.002 ***	0.003 ***	0.007 ***
		(2.73)	(2.83)	(3.34)
SI_{t-1}^{US}		0.007 ***	0.008 ***	0.011 ***
		(3.39)	(3.41)	(3.02)
Observations	175	174	174	174
Adjusted R-squared	-0.0035	0.1571	0.1296	0.1156
S.E. of regression	0.0235	0.0216	0.0219	0.0221

Panel B: IRP model

	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Intercept	2.86E-04	1.95E-04	2.02E-04	4.66E-04
	(0.14)	(0.10)	(0.10)	(0.23)
$i_{t-1} - i^{*}_{t-1}$	4.55E-05	1.15E-04	7.28E-05	1.07E-04
	(0.04)	(0.11)	(0.07)	(0.10)
SI^{Japan}_{t-1}		2.44E-03 ***	3.35E-03 ***	6.97E-03 ***
		(2.74)	(2.84)	(3.34)
SI^{US}_{t-1}		7.32E-03 ***	8.19E-03 ***	1.06E-02 ***
		(3.40)	(3.41)	(3.04)
Observations	175	174	174	174
Adjusted R-squared	-0.0058	0.1560	0.1283	0.1144
S.E. of regression	0.0235	0.0216	0.0219	0.0221

Panel C: Monetary model

	Base model	SI(30, -1)	SI(20, -1)	SI(10, -1)
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Intercept	0.00863	0.006129	0.006384	0.006629
	(0.40)	(0.31)	(0.32)	(0.33)
$m_{t-1} - m^*_{t-1} - \lambda(g_{t-1} - g^*_{t-1}) - s_{t-1}$	0.00227	0.001637	0.001692	0.001696
	(0.39)	(0.31)	(0.31)	(0.31)
SI^{Japan}_{t-1}		0.002439 ***	0.003339 ***	0.006965 ***
		(2.74)	(2.83)	(3.34)
SI^{US}_{t-1}		0.007311 ***	0.008191 ***	0.010583 ***
		(3.40)	(3.41)	(3.03)
Observations	175	174	174	174
Adjusted R-squared	-0.0049	0.1564	0.1288	0.1149
S.E. of regression	0.0235	0.0216	0.0219	0.0221

Table 4: MSPE and CW/DOC results for the AR models

Panels A to C show the MSPEs and t-statistics of the CW/DOC tests for the AR models and SI(30, -1) to SI(10, -1), respectively. SI(30, -1), SI(20, -1), and SI(10, -1) stand for the 30-word SI with a lag of 1, 20-word SI, and 10-word SI, respectively. The improvement in the MSPE is defined by the percentage decrease in the MSPE compared with the base model, and all the positive values are shown in bold for clarity. For the CW/DOC tests, *, **, and *** denote the 10%, 5%, and 1% levels of significance, respectively.

Panel A: AR model and SI(30, -1)

	AR(1)	**AR(1)+SI(30, -1)**	**AR(1)+SI(30, -1)+SI(30, -2)**	**AR(2)**	**AR(2)+SI(30, -1)**	**AR(2)+SI(30, -1)+SI(30, -2)**
MSPE (e-04)	4.953	4.938	4.920	4.906	4.818	4.967
Improvement (vs. RW)	**6.09%**	**6.38%**	**6.71%**	**6.99%**	**8.65%**	**5.83%**
Improvement (vs. AR(1))	—	**0.31%**	**0.65%**	—	**1.77%**	-1.23%
CW and DOC tests						
CW-stat (vs RW)	2.175**	2.111**	2.217**	2.311**	2.401***	2.147**
CW-stat (vs AR(1))	—	0.991	1.089	—	1.033	0.886
DOC-stat	1.054	1.476*	1.476*	1.054	0.632	1.897**

Panel B: AR model and SI(20, -1)

	AR(1)	**AR(1)+SI(20, -1)**	**AR(1)+SI(20, -1)+SI(20, -2)**	**AR(2)**	**AR(2)+SI(20, -1)**	**AR(2)+SI(20, -1)+SI(20, -2)**
MSPE (e-04)	4.953	5.208	5.184	4.906	4.848	5.233
Improvement (vs. RW)	**6.09%**	**1.26%**	**1.71%**	**6.99%**	**8.08%**	**0.77%**
Improvement (vs. AR(1))	—	-5.15%	-4.67%	—	**1.16%**	-6.62%
CW and DOC tests						
CW-stat (vs RW)	2.175**	1.687**	1.867**	2.311**	2.336***	1.806**
CW-stat (vs AR(1))	—	0.446	0.588	—	0.850	0.392
DOC-stat	1.054	0.843	0.422	1.054	0.632	1.054

Panel C: AR model and SI(10, -1)

	AR(1)	**AR(1)+SI(10, -1)**	**AR(1)+SI(10, -1)+SI(10, -2)**	**AR(2)**	**AR(2)+SI(10, -1)**	**AR(2)+SI(10, -1)+SI(10, -2)**
MSPE (e-04)	4.953	5.201	5.330	4.906	4.943	5.344
Improvement (vs. RW)	**6.09%**	**1.39%**	-1.06%	**6.99%**	**6.27%**	-1.32%
Improvement (vs. AR(1))	—	-5.01%	-7.61%	—	-0.76%	-8.85%
CW and DOC tests						
CW-stat (vs RW)	2.175**	1.872**	1.692**	2.311**	2.200**	1.687**
CW-stat (vs AR(1))	—	0.729	0.515	—	0.709	0.420
DOC-stat	1.054	1.476*	1.054	1.054	0.843	0.422

Table 5: Regression results for the AR model

The regression results for the AR(1) and AR(2) models are shown in Panels A and B, respectively. The independent variable is the log difference in monthly USD/JPY returns, and DOLLAR_YEN(-1) refers to the one-month lag of the independent variable.

Note: *, **, and *** denote the 10%, 5%, and 1% levels of significance, respectively.

Panel A: AR(1) model

	AR(1)	AR(1)+SI(30, -1)	AR(1)+SI(30, -1) +SI(30, -2)	AR(1)+SI(20, -1)	AR(1)+SI(20, -1) +SI(20, -2)	AR(1)+SI(10, -1)	AR(1)+SI(10, -1) SI(10, -2)
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Intercept	1.77E-04	1.30E-05	7.76E-05	5.96E-05	4.35E-05	3.13E-04	3.21E-04
	(0.10)	(0.01)	(0.05)	(0.04)	(0.03)	(0.20)	(0.20)
DOLLAR_YEN(-1)	0.236 ***	0.293 ***	0.329 ***	0.295 ***	0.327 ***	0.313 ***	0.315 ***
	(3.18)	(4.36)	(4.43)	(4.30)	(4.45)	(4.53)	(4.30)
SI_JAPAN(-1)		3.08E-03 ***	2.52E-03 **	4.18E-03 ***	2.91E-03 *	8.72E-03 ***	8.58E-03 ***
		(3.60)	(2.25)	(3.68)	(1.88)	(4.34)	(3.11)
SI_US(-1)		6.95E-03 ***	6.44E-03 ***	8.13E-03 ***	8.00E-03 ***	1.10E-02 ***	1.10E-02 ***
		(3.40)	(2.91)	(3.57)	(3.09)	(3.34)	(3.03)
SI_JAPAN(-2)			-9.12E-04		-1.81E-03		-2.73E-04
			(-0.84)		(-1.18)		(-0.10)
SI_US(-2)			-1.31E-03		-1.31E-03		-1.47E-04
			(-0.57)		(-0.49)		(-0.04)
Observations	174	174	173	174	173	174	173
Adjusted R-squared	0.0502	0.2409	0.2363	0.2142	0.2110	0.2097	0.1973
S.E. of regression	0.0229	0.0205	0.0206	0.0208	0.0209	0.0209	0.0211

Panel B: AR(2) model

	AR(2)	**AR(2)+SI(30, -1)**	**AR(2)+SI(30, -1) +SI(30, -2)**	**AR(2)+SI(20, -1)**	**AR(2)+SI(20, -1) +SI(20, -2)**	**AR(2)+SI(10, -1)**	**AR(2)+SI(10, -1) SI(10, -2)**
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
Intercept	5.25E-05	7.31E-05	7.24E-05	7.40E-05	3.90E-05	3.15E-04	3.34E-04
	(0.03)	(0.05)	(0.05)	(0.04)	(0.02)	(0.20)	(0.21)
DOLLAR_YEN(-1)	0.207 ***	0.280 ***	0.314 ***	0.278 ***	0.312 ***	0.294 ***	0.289 ***
	(2.73)	(3.99)	(3.99)	(3.90)	(4.02)	(4.09)	(3.72)
DOLLAR_YEN(-2)	0.121	0.058	0.043	0.065	0.046	0.071	0.075
	(1.59)	(0.84)	(0.61)	(0.93)	(0.64)	(1.00)	(1.01)
SI_JAPAN(-1)		3.09E-03 ***	2.53E-03 **	4.04E-03 ***	2.92E-03 *	8.42E-03 ***	8.70E-03 ***
		(3.44)	(2.26)	(3.44)	(1.88)	(4.06)	(3.15)
SI_US(-1)		6.95E-03 ***	6.49E-03 ***	8.17E-03 ***	8.02E-03 ***	1.12E-02 ***	1.13E-02 ***
		(3.39)	(2.92)	(3.57)	(3.10)	(3.36)	(3.09)
SI_JAPAN(-2)			-7.73E-04		-1.62E-03		4.17E-04
			(-0.70)		(-1.03)		(0.15)
SI_US(-2)			-1.24E-03		-1.18E-03		3.43E-04
			(-0.53)		(-0.44)		(0.09)
Observations	173	173	173	173	173	173	173
Adjusted R-squared	0.0587	0.2382	0.2334	0.2106	0.2081	0.2068	0.1974
S.E. of regression	0.0228	0.0205	0.0206	0.0209	0.0209	0.0210	0.0211

Table 6: Performance of the SIs across all the models

This table presents the performance of the SIs across all the models used in this study. The checkmark (✓) is given if a certain index improved the respective model by either (i) reducing its MSPE or (ii) increasing the DOC value to a statistically significant level.

Note: AR(1) + 1 lag/2 lags refers to the AR(1) model with one or two lags of the SI.

	Number of Words in Index					
Models	10	20	25	30	35	40
PPP			✓	✓		
IRP			✓	✓		
Monetary			✓	✓		
AR(1)+ 1 lag	✓		✓	✓		
AR(1)+ 2 lags			✓	✓		
AR(2)+ 1 lag		✓	✓	✓		✓
AR(2)+ 2 lags				✓		