International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Big-Data Sentiment in Exchange Rate Forecasting: Evidence from Two Million Models

Khatai Abbasov[1], Gelengül Koçaslan[2]

[1]PhD Candidate, Department of Economics (in English), Istanbul University
[2]Prof. Dr., Department of Economics (in English), Istanbul University

Abstract
This study investigates the predictability of the GBP/USD exchange rate using an exceptionally large and unrestricted set of Google search-volume indicators. An automated forecasting framework estimates approximately two million sentiment-based models and evaluates performance using strictly out-of-sample cross-validated RMSE. The study tackles three central research questions: 1) whether forecast gains from search-volume indicators are confined to a small set of finely tuned specifications; 2) whether successful models rely on similar or distinct combinations of predictors; and 3) whether the resulting model clusters correspond to economically interpretable regimes. The results show that sheer number of models outperform the benchmarks. Text-similarity analysis further reveals that these gains arise from a diverse model space rather than minor variations of a single specification. Using expert-based keyword classification, the study uncovers two consistent predictor regimes: a) slow-moving macroeconomic fundamentals; b) fast-moving financial-market sentiments. This mirrors the dual structure of exchange-rate determination emphasized in both macroeconomic and microstructure theories.

Keywords: Exchange Rate Forecasting, Statistical Learning, Google Trends, Search-Volume Data, Model Clustering, Text Similarity, Jaccard Distance, Hierarchical Clustering, Silhouette Analysis

1. Introduction

The exchange rate represents the price of one currency in terms of another. Given the interconnected nature of macroeconomic policies, capital flows, trade, tourism, and broader international linkages, fluctuations in the exchange rate, whether through appreciation or depreciation, can have profound effects on both microeconomic and macroeconomic outcomes. The foreign exchange market is highly competitive and characterized by a large number of heterogeneous participants, making it inherently challenging to forecast exchange rate movements.

Economic theory highlights four fundamental principles that shape exchange rate determination: the purchasing power parity (PPP) condition, the interest-rate parity (IRP) condition, the Fisher relation, and the rational expectations hypothesis [1, 2, 3]. The PPP condition states that exchange rates adjust in the long run to offset relative inflation differentials, ensuring that the price of identical goods evolves consistently across countries. The IRP condition links exchange rate movements to cross-country interest rate differentials, implying that arbitrage forces prevent investors from earning excess returns in one currency relative to another. The Fisher effect relates nominal interest rates to inflation expectations, i.e., higher expected inflation leads to higher nominal rates. When combined with parity conditions, this implies that currencies of countries with higher expected inflation tend to depreciate. Finally, under the rational expectations hypothesis, market participants form expectations about future exchange rates consistently with available information and underlying economic models. If agents expect the domestic currency to depreciate, they will adjust their portfolios accordingly, moving the exchange rate immediately. In efficient markets, this mechanism implies that today's exchange rate already reflects expectations of future fundamentals.

The theoretical foundations of exchange rate determination are laid out in the monetary models of Frenkel, Dornbusch, Bilson, and Frankel [4-7]. While these models offer valuable insights into the sources of exchange rate volatility, they also highlight the inherent difficulty of producing accurate forecasts. Meese and Rogoff were the first to demonstrate this challenge systematically [8]. Their influential study compared the out-of-sample forecasts of several monetary models with those of a random walk and found that none of the structural models outperformed the random walk at one-, six-, and twelve-month horizons. This striking result suggests that exchange rates appear largely unresponsive to macroeconomic fundamentals in the short run, even if fundamentals may anchor long-run movements. In this sense, the exchange rate seems to behave as if disconnected from the underlying economic factors that theory identifies [8].

A growing body of recent research shows that incorporating sentiment indicators can improve exchange-rate and macroeconomic forecasts [3, 9, 10]. Market sentiment, broadly understood as the prevailing beliefs or perceptions of economic agents, can now be proxy-measured using high-frequency online search data. Google's search volume indices, in particular, have become widely used in empirical work due to their accessibility and their ability to capture shifts in public attention [11-20]. A key challenge in this literature is determining which search terms to incorporate into prediction models, given the extremely large number of potential keywords. Fan et al. address this by applying Pearson correlation tests to identify the six search terms most strongly associated with Taiwan's stock market index [18]. Da et al. adopt a different approach by selecting the 30 search terms with the largest negative t-statistics in a backward-rolling regression of a search-data index on stock market returns, capturing fear-related searches that tend to precede market declines [16]. These selected terms are then aggregated into the Financial and Economic Attitudes Revealed by Search (FEARS) index, which is shown to predict U.S. stock market dynamics effectively.

The foreign exchange literature also exploits Google Trends data extensively. Masuda and Takeda, for example, build on the methodology of Da et al. [16] and construct FEARS indices for Japan and the United States to capture market sentiment in forecasting the JPY/USD exchange rate [3]. To address the issue of selecting relevant search terms from a large and flexible set of possibilities, Chojnowski and Dybka [9] and Yaganti and Manpuria [10] apply principal component analysis to reduce the dimensionality of the search-volume data to a single dominant factor. These factors are then used to construct price, credit, and financial market sentiment indices, which are incorporated into vector autoregressive models to forecast the PLN/EUR and INR/USD rates. Other studies take a more direct approach by modeling the public's attention to the exchange rate itself, relying on search terms such as "MXN/USD," "dollar rate," or "Mexican peso" as proxies for short-run trading sentiment and exchange-rate pressure [21, 22]. Bulut offers yet another perspective by examining whether macroeconomic models augmented with Google search data can outperform the random walk or traditional structural models [2]. His motivation is that official macroeconomic data, although theoretically informative, are released with substantial delays, whereas search behavior may reflect real-time perceptions of prices, income, and liquidity. Across all these studies, the inclusion of sentiment indices derived from Google search data leads to meaningful improvements in forecasting accuracy.

This study contributes to this expanding literature on sentiment-based exchange-rate forecasting but departs from the standard practice in an important methodological dimension. Rather than imposing ex-ante restrictions on keyword selection or reducing dimensionality prior to forecasting, this paper incorporates the full universe of roughly 25,000 Google search-volume series as a high-dimensional predictor space. The first methodological contribution is therefore to evaluate exchange-rate predictability using an exceptionally large and unrestricted information set. Because of the scale of this predictor space, variable selection is performed entirely through an automated forecasting framework that estimates approximately two million models.

This motivates the first research question:

- Are forecast improvements concentrated in a small number finely-tuned models, or do they arise systematically across a broad model space?

This distinction is central because if improvements emerge across thousands of independent specifications, then sentiment signals are robust and broadly informative. The findings confirm that the gains are indeed systematic, with a large share of models outperforming the random walk and autoregressive benchmarks in terms of root mean squared error (RMSE).

The second research question concerns the structure of these successful models:

- Do sentiment-based models achieve superior performance by relying on similar sets of variables, or through distinct combinations of search-volume predictors?

To examine this, the study employs text-similarity tools, such as Jaccard distances, hierarchical clustering, and silhouette widths, to quantify the overlap in variable sets across models. This constitutes the second major contribution, as such a large-scale similarity assessment has not, to the best of our knowledge, been conducted in the foreign exchange (FX) forecasting literature.

The third research question asks:

- Do model clusters formed using text-similarity distances correspond to economically interpretable regimes?

To explore this, keyword groups within each cluster are classified into economically meaningful subcategories using expert judgment. Two regimes emerge consistently. The first represents slow-moving fundamentals, such as prices, output conditions, and other real-economic indicators. The second regime is dominated by financial-market variables, such as interest-rate conditions, liquidity signals, and short-run trading sentiment. The interpretation of this second regime is strengthened by advances in the microstructure theory that stresses the importance of short-run trading behavior in exchange-rate determination. Evans and Lyons show that order flow, the imbalance between buyer- and seller-initiated FX trades, contains substantial information about future exchange rate movements because it aggregates dispersed private signals that are not captured by macroeconomic variables [23]. Their later work demonstrates that even public macroeconomic announcements affect exchange rates largely through the order flow [24]. These insights help explain why search-volume data reflecting financial-market attention and trading sentiment capture short-run variation in exchange rates.

Although the classification derived from search-volume data is necessarily heuristic, it aligns closely with the long-established duality in exchange-rate determination: short-run movements are shaped by financial-market dynamics, while long-run adjustments reflect gradually evolving macroeconomic

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

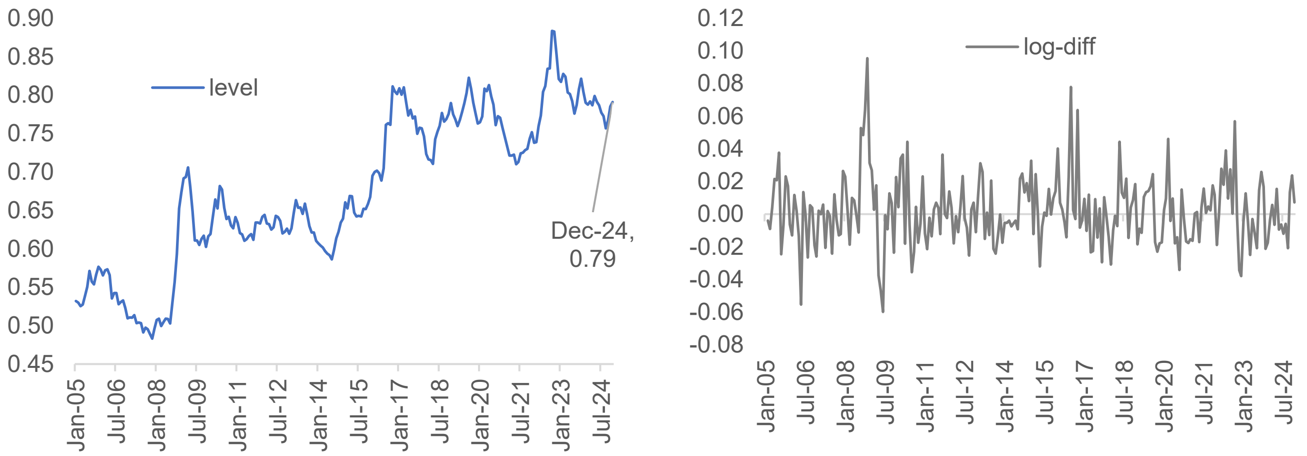
fundamentals. Linking data-driven clusters to underlying economic mechanisms therefore constitutes the third contribution of the paper.

While these research questions are introduced here, the subsequent sections analyze each in depth and provide rigorous empirical answers

2. Data

The exchange rate in this research is the monthly rate of British pound to US dollars. The data covers the periods from January 2005 to December 2024. Although a twenty-year window of information is sufficient, it should noted that taking this particular timeframe is directly linked to the availability of Google data, which forms the main component of high-frequency alternative dataset here. The sources of the utilized series along with description of applied transformations on them are available in Appendix 1.

Figure 1: GBP/USD rate, 01/2005-12/2024, monthly

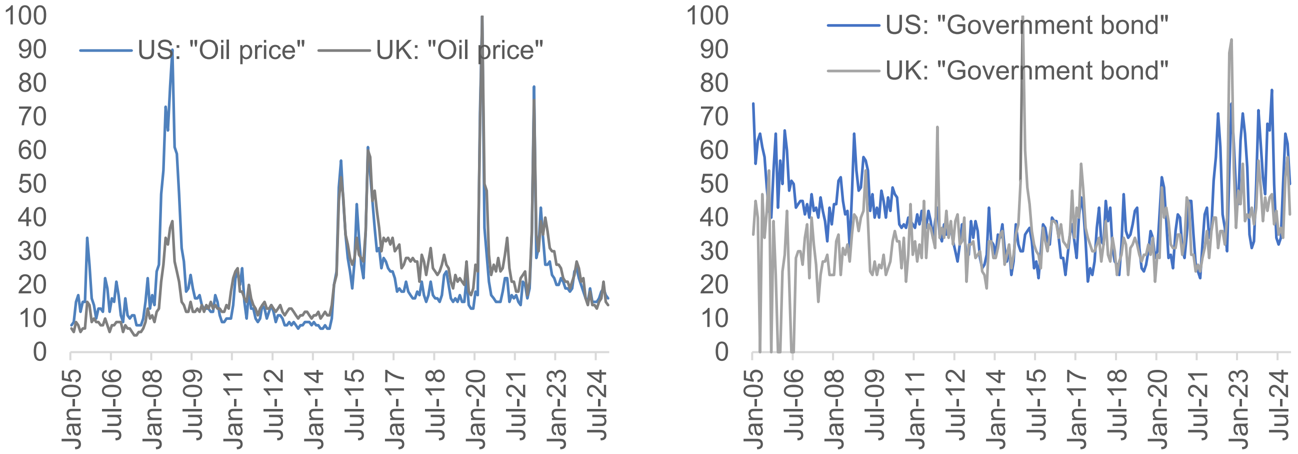


Source: Federal Reserve Bank of St. Louis

The explanatory data are built by means of Internet search frequency thanks to Google. The data are updated on weekly base and available for any country commencing from 2004 on the webpage of Google Trends. The main problem here is to find a time-invariant words that are relevant in the context [25]. Therefore, this article exploits a broad set of potentially relevant words that may affect the variance of the exchange rate. The set of used keywords is available with alphabetical order in Appendix 2. To create this list, authors initially scanned the literature for all used keywords in the stock or foreign exchange market analysis via search volume data of Google. Then, they skim through the online and offline materials for all relevant financial terms, abbreviations, stock market and foreign exchange market terminologies. Furthermore, authors incorporate the words that are the names of the famous financial webpages that provide financial news or data.

There are 505 keywords that are used to collect data from the webpage of the provider Trends. Illustration for how to downloading these search volume indices is available in Appendix 2. As the focus is on the exchange rate of British pound against the US dollar, data is gathered for each country. Therefore, total number of variables reaches to 1010. Exemplary plots for some search volume indices are shared in Figure 2. As the platform normalizes the data by dividing each observation by the total search volume for the corresponding region and time period, and then scales it from 0 to 100, each index has a maximum value of 100. The minimum value varies across series. These numerical values reflect the digital representation of people's interest in a given currency during a specific period and location. For example, in both locations, the highest interest in "oil price" is reflected by the number 100 in April 2020, which is not surprising given the crude oil futures price fell below zero for the first time ever. Or a spike in searches for "government bond" in the UK is observed around January 2015 when the UK government bond yields hit record lows. So, both investors (trying to understand market risk) and ordinary people (seeing bizarre headlines) searched for explanations on Google.

Figure 2: Google Trends Search Volume Indices, 01/2005-12/2024, monthly



Note: This figure provides demonstrative plots for few series as there are 1010 indices in total.
Source: Google Trends

Trends in the series of exchange rate and Google Trends indicate that the variables may not be stationary. Given the sheer volume of the dataset, the Augmented Dickey-Fuller (ADF) and Phillips-Perron unit root tests, along with the Kwiatkowski-Phillips-Schmidt–Shin (KPSS) stationarity test, are automated. Series are tagged as non-stationary if at least one of these tests fails at the 5% significance level. Afterwards, stationarity transformations are applied using either first differences or log-differences (first differencing the natural logarithm of the series). Although most variables, including the exchange rate, become stationary after the transformation, ten of the search volume series remain non-stationary in at least one of the aforementioned tests, and two of the series has no enough data to utilize. These 12 variables are therefore excluded from the final set of inputs used in the models.

Table 1: Summary of Variable Stationarity Status

	Stationary after Transformation	**Nonstationary after Transformation**	**Total Number of Variables**
Level Nonstationary	903	10	913
First-difference	177	1	178
Log-difference	726	9	735
Level Stationary	-	-	96
Not Enough Data	-	-	2
Grand Total	-	-	**1011**

Note: Total number of variables from Google Trends is 1010 in addition to exchange rate of GBP/USD. The log-difference transformation is obtained by first differencing the natural logarithm of the series.
Source: own visualization

3. Methodology

This article employs an automated forecasting methodology that standardizes model estimation, evaluation, and comparison across both autoregressive and regression-based approaches. The framework combines lag-structured data preparation, recursive model estimation under rolling-origin and rolling-window designs, and out-of-sample projections. All variables are first transformed into a structured lag matrix, i.e., an automated lag-generation procedure constructs lagged versions of each series from lag 0 up to 24 months. This transformation enables both univariate and multivariate model setups and provides a coherent data foundation for all subsequent analyses. Forecasts are produced out-of-sample after each automated estimation, ensuring that each prediction relies strictly on information available at the time of the forecast. This framework yields a sequence of forecast errors that reflect genuine real-time performance rather than in-sample fit. Moreover, the framework utilizes rolling-origin and rolling-window designs during the estimations.

- **Rolling-origin forecasting:** The training sample expands over time. After an initial estimation window of 100 months, the model is re-estimated at each new observation while forecasts are generated for the next one period. In other words, each model is run with the time window of $100 + \mathrm{i}$, where the value for $\mathrm{i} = \{0,1,2,3, \ldots .\mathrm{n}\}$ goes up to n where the $100 + \mathrm{n} = \mathrm{N}$, which is the total number of all available observations in that series. Thus, each model creates $\mathrm{n} + 1$ cross-validated out of sample predictions and prediction errors.
- **Rolling-window forecasting:** The estimation window length remains fixed. As time advances, the oldest observation is dropped, and the newest is added. In other words, each model is run with the time window of 100 months, which is equal to the initial 100 observations in the first estimation but the last 100 observations in the final estimation for that series. This design allows the analysis of model stability when structural breaks or parameter drift may occur. Yet, each model creates $\mathrm{N} - 100 + 1$ cross validated out of sample predictions and prediction errors, where the N is he total number of all available observations in that series.

As the focus is on the prediction power of Google search volume sentiments on the exchange rate of GBP to USD, the performance of models is measured and contrasted via the RMSE, which is a widely used metric in similar forecasting and econometric analyses. Mathematically, the RMSE calculates the square root of the mean of the squared differences between predicted and actual values of the exchange rate, i.e., prediction errors that are obtained during the rolling-origin and rolling-window forecasting. Lower RMSE values indicate higher predictive accuracy.

Given the high dimensionality of the input data, around 25,000 variables, the regression component employs a stepwise forward learning algorithm for variable selection. Starting with a univariate model, additional regressors are introduced iteratively based on their improvement in cross-validated out-of-sample RMSE. The process continues until a predefined level of parsimony is reached, corresponding to a maximum of six-variable models, in order to avoid overfitting and reduce estimation costs. To account for the similar nature of the series and the potential issue of multicollinearity in the models, variables exhibiting correlations above 90% are excluded from the forward learning process. All in all, approximately 2,000,000 model-driven RMSE values are generated for each of the cross-validated rolling-origin and rolling-window forecasting designs. Table 2 presents a schematic representation of this process. It is noteworthy to reiterate that variable selection is performed separately in the rolling-origin and rolling-window designs, meaning that the variables included in the models may differ from one another due to the different treatment of the estimation window.

Once the out-of-sample RMSE values are obtained for each model in the stepwise forward learning design, the results are compared both internally and against those from a random walk without drift, as well as against rolling-origin and rolling-window cross-validated AR(p) models with lag orders ranging from 1 to 24 months. This comparison is in line with [8] and aims to assess whether the models incorporating Google search sentiment outperform the random walk or the AR(p) benchmarks. The results of all these models are presented in the next section.

Table 2: Automated Stepwise Forward Selection Schema

	Model	Note
1st Stage	$Y \text{ reg } X_1^i$	X_1^i is one of nearly 25,000 lagged variables
2nd Stage	$Y \text{ reg } X_1^i \,;\, X_2^i$	X_1^i and X_2^i are the combination of the lowest 700 RMSE producing variables at 1st stage Given 90% correlation threshold to avoid multicol-linearity, more than 240,000 models are created at 2nd stage
3rd Stage	$Y \text{ reg } X_1^i \,;\, X_2^i \,;\, X_3^i$	X_1^i and X_2^i being the variables of the lowest 700 RMSE producing models at 2st stage are combined with X_3^i, which is one of the lowest 700 RMSE producing variables at 1st stage Given 90% correlation threshold to avoid multicol-linearity, more than 400,000 models are created at 3rd stage
4th Stage **5th Stage**	$Y \text{ reg } X_1^i \,;\, X_2^i \,;\, X_3^i$ $Y \text{ reg } X_1^i \,;\, X_2^i \,;\, X_3^i$	…
6th Stage	$Y \text{ reg } X_1^i \,;\, X_2^i \,;\, X_3^i \,;\, X_4^i \,;\, X_5^i \,;\, X_6^i$	$X_1^i \,;\, X_2^i \,;\, X_3^i \,;\, X_4^i \,;\, X_5^i$ being the variables of the lowest 700 RMSE producing models at 5th stage are combined with X_6^i, which is one of the lowest 700 RMSE producing variables at 1st stage Given 90% correlation threshold to avoid multicol-linearity, more than 400,000 models are created at 6th stage
Total		**Around 2,000,000 models**

Note: The process is limited to the top-performing 700 variables at the first stage to prevent the maximum model count from exceeding 500,000 at each stage. While reducing computation time, this approach also helps avoid potential “garbage in, garbage out” effects. The term reg in the table stands for regression, i.e., Y reg X means $Y = \text{intercept} + \beta_1 X$. The RMSE of each model is computed based on cross-validated rolling-origin and rolling-window 1-step-ahead prediction errors.

Source: own visualization

Given the large number of estimated models, and the vast amount of potential input variables, it is possible that many specifications of interest are just slight variants of one another. To address this concern, the degree of similarity among the best-performing models is further examined by applying text-based clustering techniques. Specifically, the Jaccard distance is utilized to quantify how distinct the models are with respect to their selected variables. This allows for an assessment of whether the leading specifications represent genuinely different variable combinations or minor replications of the same underlying structure. Mathematically, let X and Z denote the right-hand-side variable sets of two different models, and compute the Jaccard distance (JD) as follows:

$$\mathrm{JD}\,(\mathrm{X},\mathrm{Z}) = 1 - \frac{|\mathrm{X} \cap \mathrm{Z}|}{|\mathrm{X} \cup \mathrm{Z}|}$$

The Jaccard distance ranges from 0 to 1, where a value of 0 indicates that the models use exactly the same predictors, whereas a value of 1 implies they share none. Thus, a diverse set of high-performing models would suggest that multiple, distinct configurations of Google search-based sentiments can successfully forecast the exchange rate. In addition to pairwise Jaccard distances, authors also apply hierarchical clustering to group the top-performing models based on their similarity. Hierarchical clustering uses the Jaccard distance as the input and merges models into clusters, starting from the most similar pairs. The result is a tree-like structure (dendrogram) that shows which models are closest to each other and how model groups form as the similarity threshold increases. This helps to analyze whether the best-performing models concentrate around distinct clusters of variables [26, 27].

$$\text{Silhouette Width (i)} = \frac{\mathrm{b(i)} - \mathrm{a(i)}}{\max\,(\mathrm{a(i)}, \mathrm{b(i)})}$$

The separation of the model clusters is evaluated using silhouette widths, which is a standard metric to assess cluster quality. For each model i, $a(i)$ = average Jaccard distance between model i and models in its own cluster, and $b(i)$ = average Jaccard distance between model i and models in the nearest other cluster. Silhouette width ranges from -1 to +1 and measures how well a model fits its assigned cluster compared to the next-closest cluster. A value close to +1 means that the model is well classified, whereas a negative value suggests that the model may be misclassified or that cluster is poorly defined. In practice, positive silhouette peaks in between 0.3 and 0.6 are interpreted as evidence of good clustering [28].

4. Results

The RMSE results for the AR models up to 24-month reveal that all AR models, both in rolling origin and rolling window designs, perform better than the random walk (RW) benchmark, with the AR(1) model producing the lowest RMSE. More strikingly, models incorporating Google Search Volume Sentiment demonstrate even better performance than the AR(1) and RW benchmarks. The lowest RMSEs achieved among these sentiment models are reported in Table 3.

Table 3: Lowest RMSE Across Models

	RMSE Value
Random Walk	0.02502
AR(1) in Rolling Window	0.01903
AR(1) in Rolling Origin	0.01887
Best Sentiment Model in Rolling Origin	0.01490
Best Sentiment Model in Rolling Window	0.01447

Note: Random Walk without drift. AR(1) model is selected as benchmark because it has the best accuracy among all AR models with lag orders up to 24 months. For the AR models, the rolling-origin and rolling-window designs indicate that the reported RMSE values are derived from cross-validated one-step-ahead forecasts, where the estimation window is expanding or fixed, respectively. The Best Sentiment Model refers to the specification that achieves the lowest RMSE among all cross-validated stepwise forward learning models that utilize Google Search Volume Indices.
Source: own estimation

A natural question that arises is whether it is only a single model (or a small number finely-tuned models) that outperform the AR(1) and RW benchmarks. In other words, do the sentiment models systematically deliver lower RMSE values than the benchmarks, or are the observed improvements solely coincidental? Figure 3 reveals that the majority of the sentiment models indeed perform better than their benchmark counterparts, even though they had not been programmed to do so. Indeed all model space could have been worse than the benchmarks despite having the lowest RMSE within programed design. In the rolling-origin design, approximately 91% of the two million models outperform the AR(1) benchmark, whereas in the rolling-window design this share increases to about 98.5%. Moreover, nearly the entire set of sentiment models exhibits better accuracy than the random walk model. To evaluate whether the RMSE improvements of sentiment-based models over the AR(1) benchmarks are statistically meaningful, a nonparametric bootstrap is applied to the distribution of RMSE differences across all cross-validation folds. Using 10,000 bootstrap replications, the 95% confidence interval for the mean RMSE difference is tight and strictly negative both in rolling-origin design $(-0.001802, -0.001799)$ and rolling-window design $(-0.001935, -0.001932)$. A simple paired t-test leads to the same conclusion, yielding highly significant statistics for both rolling-origin $(\mathrm{t} = -2410, \mathrm{p.val} < 0.001)$ and rolling-window $(\mathrm{t} = -2468, \mathrm{p.val} < 0.001)$. Taken together, these results confirm that the observed improvement is highly stable and not driven by sampling variation. Therefore, models that incorporate Google search volume sentiments have consistent systematic gains rather than coincidental.

Table 4: Input Variables of Top Five Sentiment Models

	Rolling Origin Design	**Rolling Window Design**
Model 1	$=$ Oil_price_in_US_lag0 $+$ Pound_account_in_UK_lag0 $+$ Schengen_in_UK_lag4 $+$ Initial_public_offering_in_UK_lag8 $+$ Swiss_Franc_in_US_lag18 $+$ Charity_in_UK_lag20	$=$ Dollar_Rate_in_US_lag0 $+$ Government_bond_in_UK_lag0 $+$ EU_in_UK_lag1 $+$ GBP_in_US_lag4 $+$ Initial_public_offering_in_UK_lag8 $+$ Dollars_in_UK_lag10

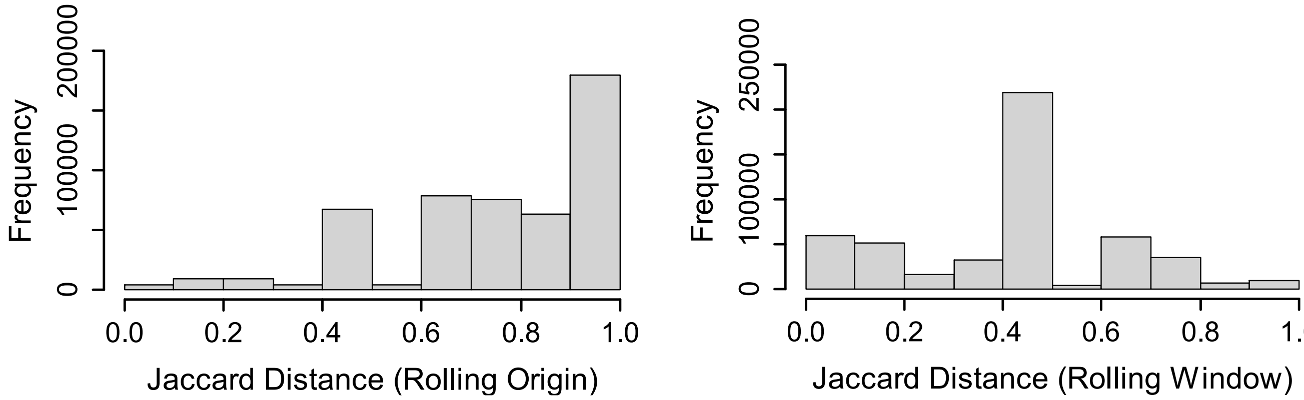
Model 2	= Dollar_Rate_in_US_lag0 + Government_bond_in_UK_lag0 + EU_in_UK_lag1 + Prediction_in_UK_lag3 + Initial_public_offering_in_UK_lag8 + EUR_in_US_lag21	= Dollar_Rate_in_US_lag0 + Government_bond_in_UK_lag0 + EU_in_UK_lag1 + GBP_in_US_lag4 + Chevron_in_US_lag6 + Initial_public_offering_in_UK_lag8
Model 3	= Oil_price_in_US_lag0 + Bank_of_England_in_UK_lag0 + Schengen_in_UK_lag4 + Initial_public_offering_in_UK_lag8 + Swiss_Franc_in_US_lag18 + Charity_in_UK_lag20	= Dollar_Rate_in_US_lag0 + Government_bond_in_UK_lag0 + EU_in_UK_lag1 + GBP_in_US_lag4 + Volatility_in_US_lag6 + Initial_public_offering_in_UK_lag8

Note: The output variable for all models is the GBP/USD exchange rate. These models yield the lowest RMSE, computed from cross-validated rolling-origin and rolling-window 1-step-ahead prediction errors.
Source: own visualization

As an illustration, the right-hand-side variables of the top three performing models for both the rolling-origin and rolling-window designs are presented in Table 4. The dependent variable for all these models is the exchange rate of the British pound against the US dollar. The top-performing models in the table show both similarities and differences in their input variables. For example, under the rolling-origin design, Model 1 and Model 3 share most of the same variables, differing only by the inclusion of Pound_account_in_UK in Model 1 and Bank_of_England_in_UK in Model 3. In contrast, Model 2 is genuinely different as it shares at most one variable with others. Under the rolling-window design, all three models are just slight variants of one another as they have four variables in common but differ in one.

Many variables rise to the top simply because the stepwise forward-learning algorithm reuses them. This raises an important question: do sentiment-based models achieve superior performance by relying on similar sets of variables, or through distinct combinations of search-volume predictors? Given the vast number of potential input variables and the large number of estimated models, it is not feasible to address this question manually. For this purpose, authors turn to text similarity measure and take advantage of the Jaccard distance, which helps to quantify whether the leading specifications are genuinely distinct or simply minor replications of the same underlying structure. To illustrate with an example, if two models share 4 variables out of 7 total variables used in either one, the Jaccard distance will be $1 - 4/7 = 0.429$. In general, distances in the range of $0.6 - 1.0$ suggest that models are quite distinct from one another, sharing very few variables. Values near $0 - 0.3$ indicate groups of highly similar models, i.e., specifications use almost the same set of variables.

Figure 4: Model Similarity Distributions

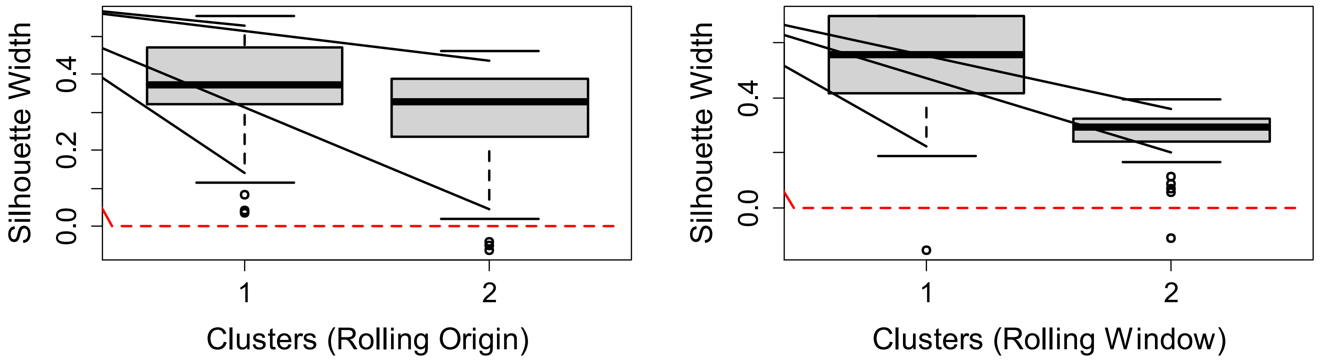


Note: This figure displays Jaccard distance distributions for the top 1,000 models with the lowest RMSE under rolling-origin and rolling-window one-step-ahead cross-validated forecasts. The Jaccard distance reflects variation in the independent variables across models, indicating how similar (values close to 0 on the x-axis) or how distinct (values close to 1 on the x-axis) the right-hand-side variable sets are.
Source: own estimation

Authors take the top 1,000 performing models and examine how distinct they are in terms of their input variables. The choice of 1,000 is deliberate as it is large enough to capture meaningful variation across models, yet small enough to focus attention on the highest RMSE part of the performance distribution. The Jaccard distance is computed for the rolling-origin and rolling-window designs, and are plotted in Figure 4. What stands out is that, under the rolling-origin design, many models are clearly distinct from one another, with the majority of Jaccard distances exceeding 0.6. In contrast, the distances in the rolling-window design are generally lower. This is largely due to the fixed-length window, which forces the model to fit short-term time-local correlations, not global structure. A rolling window uses only a small recent subset of data (100 monthly observations in this article) that reduces the model's ability to exploit long-run dynamics, structural breaks, and macroeconomic relationships that changes gradually over time. In other words, narrow window width destroys long-run relationships and results in fewer "good" variables that are noisy and fast-developing candidates. As a result, the top-performing rolling-window models tend to be more similar to one another, producing less distinct variable regimes compared with the rolling-origin design.

Next hierarchical clustering is used to group the top-performing 1,000 models based on their similarity. Hierarchical clustering takes the Jaccard distance as input and iteratively merges models into clusters, beginning with the most similar pairs. Examining the cluster-quality metric of silhouette widths in Figure 5, two clusters are identified in the rolling-origin design. The first cluster has silhouette value mean/median around 0.4, suggesting a statistically meaningful structure. Although the second cluster is less clear than the first, its median silhouette value is above 0.3, indicating that it can still be treated as a distinct cluster. For the rolling-window design, the results signal only one meaningful cluster, with mean and median silhouette values above 0.5. This reflects a more similar set of variables across the top-performing models. The second cluster containing only 27 models, has silhouette values below 0.3 and a less distinct set of variables, and therefore cannot be considered a well-defined cluster as initially expected.

Figure 5: Cluster Quality of Model Groups



Note: This figure displays the silhouette widths of clusters formed from the Jaccard distances of the top 1,000 models for both the rolling-origin and rolling-window designs. Silhouette values between 0.3 and 0.6 indicate meaningful cluster structure, values near zero suggest weak or unclear clustering, and negative values point to poorly defined or misclassified clusters.
Source: own estimation

Do model clusters formed using text-similarity distances correspond to economically interpretable regimes? To examine this, authors classify the variables into subgroups based on their own expert judgement. In doing so, we consider the potential motivations behind each Google search term and how it might relate to movements or variance in the exchange rate. The classification of the most frequently selected variables within each cluster is provided in Appendix 3. Based on their frequencies, the resulting subcategories, along with their total frequencies and corresponding shares, are summarized in Table 5.

Table 5: Model Input Subgroups and Their Relative Importance Across Clusters

	R. Origin: Cluster 1 Frequency, Share	**R. Origin: Cluster 2** Frequency, Share	**R. Window: Cluster** Frequency, Share
Output	670, 24%	192, 7%	78, 2%
Price	526, 19%	127, 5%	66, 1%
Rate	727, 26%	1615, 62%	3662, 74%
Trade	333, 12%	541, 21%	1010, 21%
Risk	84, 3%	130, 5%	104, 2%
Safe Haven	487, 17%	-	-

Note: Summary of the economic subcategories assigned to the input variables in Appendix 3 via expert judgement. The reported frequencies capture how often variables within each subcategory appear across the top-performing models. Shares are obtained by dividing each subcategory's frequency by the total frequency reported for the given cluster. Rounding error in shares.
Source: own estimation and classification

In Cluster 1, the exchange rate responds primarily to price levels, real-economy shocks, commodity conditions, trade activity, inflation signals, safe-haven flows, and household sentiment. The regime displays a strong presence of commodity-related variables (oil, crude oil, gold), real-sector indicators (mining, charity, housing, shortages), and structural macroeconomic institutions (such as the Bank of

England and the Euro Area). Financial risk terms are relatively limited, while safe-haven activity appears consistent with globally inflationary conditions and heightened uncertainty, though without evidence of systemic panic. Overall, Regime 1 is associated by slow-moving macroeconomic fundamentals and relative price structures. This pattern aligns with purchasing-power-parity adjustments, commodity-price pass-through, real shocks, output and inflation dynamics, and moderate safe-haven repositioning.

In Cluster 2, by contrast, the exchange rate is predicted overwhelmingly by interest rates, yields, conversion activity, and market microstructure. The regime is characterized by a high frequency of terms associated with financial-market operations, including arbitrage, over-the-counter trading, FX platforms, traders, London Interbank Offered Rate (LIBOR), futures, margins, daily FX flows, and interbank markets. This indicates a financial-market-dominated environment in which exchange rates adjust to interest-rate differentials (IRP), capital flows, expected returns, liquidity conditions, arbitrage pressures, and speculative activity. Real-sector variables play a minimal role, reflecting the classic signature of an asset-market exchange rate regime. Taken together, Regime 1 corresponds to a real-monetary exchange rate framework, whereas Regime 2 corresponds to an asset-market microstructure or interest-rate-parity-associated regime.

Analysis of the rolling-window design indicates that its single dominant cluster largely aligns with the financial-market-driven family of variables associated with Regime 2. This pattern is consistent with the notion that, over short horizons, exchange rate movements respond primarily to interest-rate differentials, liquidity conditions, risk premia, and market microstructure forces, whereas slowly evolving fundamentals such as output, price levels, or commodity terms of trade contribute little predictive power.

5. Conclusion

This article investigates the predictability of the GBP/USD exchange rate using an exceptionally large set of Google search-volume indicators over the period January 2005 to December 2024. To ensure a meticulous and transparent evaluation, an automated forecasting framework was developed that standardizes model estimation, comparison, and forecast assessment across millions of specifications. The framework incorporates up to 24 lags of each variable, uses strictly out-of-sample rolling-origin and rolling-window designs, and evaluates predictive accuracy through recursive estimation with cross-validated RMSE. All forecasts rely exclusively on information that would have been available at the time predictions were made.

The motivation stemmed from longstanding evidence that traditional macroeconomic models struggle to outperform naive benchmarks out-of-sample [8]. At the same time, a growing empirical literature has demonstrated that information-search behavior can contain predictive signals not captured by conventional macroeconomic variables [2, 3, 9, 10, 16, 21, 22]. Against this backdrop, the present work engaged with three central research questions concerning

- the existence and breadth of forecast gains
- the composition and similarity of successful models
- the economic interpretability of resulting predictor clusters.

The analysis demonstrates that incorporating Google search-volume data meaningfully improves the forecasting performance for GBP/USD relative to both the random-walk benchmark and autoregressive models. These gains are not confined to a handful of isolated, finely tuned specifications. Instead, forecast improvements emerge systematically across a very broad model space: thousands of independently estimated models deliver lower RMSEs, and in some estimation designs, the fraction of models that outperforms auto-regressive benchmarks exceeds 90%. This provides strong evidence that sentiment-based information embedded in large collections of search terms carries stable, reproducible predictive content. The finding directly addresses the first research question and constitutes the study's first methodological contribution: evaluating exchange-rate predictability using an exceptionally large, unrestricted, high-dimensional information set rather than a manually curated subset of search terms.

The study also examines whether successful models rely on similar or different sets of predictors. Using text-similarity metrics, such as Jaccard distances, hierarchical clustering, and silhouette widths, the analysis reveals that high-performing models are not simply minor variants of one another. Instead, they span multiple distinct combinations of keywords, indicating that predictive improvements arise from a diversified space of variable sets, not a single dominant specification. This satisfies the second research question and constitutes the second methodological contribution: implementing a large-scale structural analysis of the model space itself, which to the best of our knowledge has not previously been done in the exchange-rate forecasting literature.

Next, the data-driven clusters of predictors are linked to economic mechanisms through expert-based keyword classification. Two coherent regimes emerge. One encompasses slow-moving macroeconomic fundamentals, such as price levels, output conditions, liquidity constraints, and other real-economy indicators, that are consistent with long-run theories such as purchasing-power parity and other theoretical models. The second regime is dominated by financial-market variables, short-run trading sentiment, liquidity perceptions, and attention to currency-conversion pressures. This structure mirrors the duality documented in both traditional macro-finance theory and the microstructure literature, particularly the work of Evans and Lyons, who show that short-run exchange-rate movements are heavily shaped by trading flows and market-level information aggregation [23, 24]. This mapping provides the answer to the third research question and represents the paper's third contribution: bridging high-dimensional text-derived clusters with economically interpretable exchange-rate regimes.

Taken together, the findings offer a comprehensive picture of how sentiment expressed through online search activity is embedded in exchange-rate dynamics. Search-volume indicators contain rich and persistent information about both macroeconomic perceptions and financial-market conditions. They improve forecasting accuracy in a manner that is broad-based, structurally diverse, and economically interpretable. The methodological contributions of this research also extend beyond exchange-rate forecasting. The scalable framework developed here (combining unrestricted predictor universes and text-mining-based structural diagnostics) offers a flexible blueprint for future empirical work in macro-finance, behavioral economics, and big-data-augmented forecasting.

Despite its contributions, the study also leaves room for further exploration. A valuable extension is to apply the framework to additional currency pairs to assess whether the macro-financial sentiment duality identified here is universal or specific to GBP/USD. Another direction is to contrast the results with nonlinear and machine-learning models (LASSO regression, random forests, gradient boosting, neural networks) to determine whether more flexible functional forms extract additional information from search data.

Finally, future research could investigate how the length of the estimation window affects the balance between predictor regimes. The present analyses rely on a fixed rolling-window size of 100 observations, which already favors short-run financial and sentiment-based variables over slow-moving macroeconomic fundamentals. Varying the window length (e.g., 30, 50, 80, and beyond) would allow researchers to trace how the relative importance of microstructure-driven signals declines as the horizon expands and macroeconomic fundamentals re-enter the predictive structure. Such extensions would deepen our understanding of the horizon-dependent mechanisms through which sentiment and fundamentals interact in shaping exchange-rate dynamics.

6. Appendix 1: Description for USD to GBP FX Rate

- U.S. dollars to 1 U.K. Pound Sterling, Daily, Not Seasonally Adjusted. Monthly version of the data is calculated with an arithmetic average of the corresponding period. The data is exponentiated to -1 to interpret as British pound to 1 US dollars.
- The original data is accessible in the webpage of Federal Reserve Bank of St. Louis (Ticker: DEXUSUK)

7. Appendix 2: Description for Google Search Volume Indices

The keywords in Table 6 are used to collect data from the webpage of Google Trends as it is depicted in Figure 6. Time range is customized to start from Jan-2005 on. The retrieved data is monthly and not seasonally adjusted.

As Google scales the data on a range of 0 to 100 after normalizing it by dividing each data point by the total searches within the corresponding geography and time range, collecting multiple indices simultaneously from the same page results in relative scaling, potentially reducing the volatility of each series significantly. To mitigate such cases, the series are downloaded one by one for each country. The further challenge is that some data points have not certain value but are given as less than one. To overcome this problem, the value of 0.5 is overwritten whenever the datapoint is less than 1.

Table 6: Keywords Used to Create Google Trends Search Volume Indices

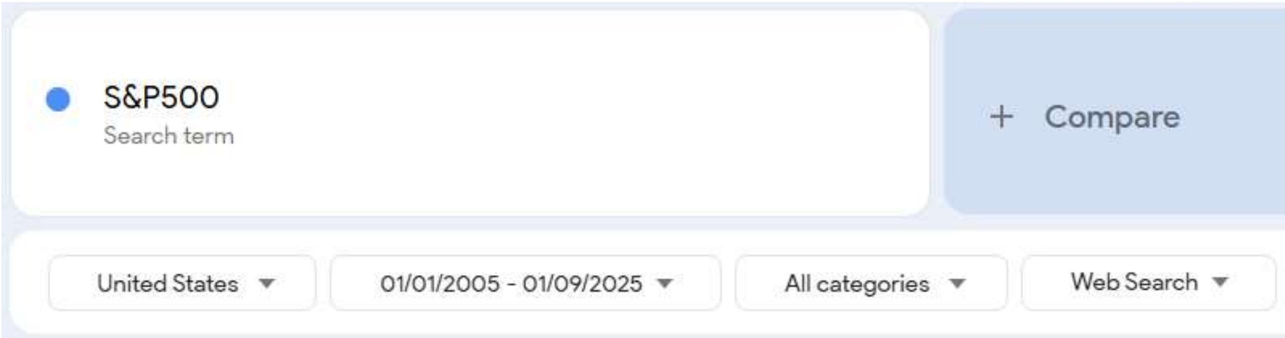
Account : Accountant : Accounting : ACER : Acquisition : America : America tax : Annual report : Annualized : Arbitrage : Army : Arts : Ask price : Associate : ASUS : ATM : AUD : Audi : Audit : Australia : Australian dollar : Bank : Bank loan : Bank of England : Banking : Bankrupt : Bankruptcy : Bargain : Base Currency : Baseball : Basketball : Battery : Bear market : Bearish : Benefits : Bid : Bid price : Bloomberg : Blue chip stocks : Bomb : Bond : Bourse : Bowling : BP : Britain : Britain tax : British companies : British : British income tax : British Pound : British Sterling : British stock : Broker : Bubble : Bull market : Bullish : BUY AND HOLD : Buy : Canadian Dollar : Cancer : CapEx : Capital gain : Capital : Capital loss : Capitalization : Car : Cargo : Cargo insurance : Cash : Cboe : Central Bank : CETA : Chance : Charity : Cheap : Chevron : CHF : China currency : China foreign exchange : China GDP : China : China income tax : Chinese companies : Chinese currency : Chinese market : CMC : Colony : Color : Computer : Conflict : Consume : Consumption : Convert currency : Cooperative : Coupon : CPI : Crash : Crawling peg : Credit : Credit Insurance : Crisis : Crude Oil : Culture : Currency Converter : Currency Market : Currency Pair : Customs : Daily forex : Daily fx : Day trading : Dealer : Debt : Default : Deficit : Deposit account : Deposit : Depression : Derivatives : Deutsche Borse : Dividend : DJIA : Dollar account : Dollar deposits : Dollar Rate : Dollar to Pound : Dollars : Domestic : Donation : Dow Jones : Dowjones : Earnings : EBIT : EBITDA : EBT : Economic risk : Economics : Economy : Effective Exchange Rate : Election : Elections : Electric car : Energy : England : England income tax : English :

Entrepreneurial : Environment : Equity financing : Equity : Estate : EU : EUR : Euro account : Euro Area : Euro : Euronext : Europe : Europe income tax : Euros : Exchange : Exchange office : Exchange rate : Exchange Rate Risk : Exchange rates : Expense : Explosion : Export : Exposure : Exxon : FCF : FED : Federal income tax : Federal Reserve : Fellowship : Finance : Financial crisis : Financial : Financial Institution : Financial lease : Financial market : Financial markets : Financial statement : Fine : Fiscal policy : Fixed currency : Fixed exchange rate : Flat : Floating currency : Floating exchange rate : Fond : Food : Ford : Forecast : Forecasting : Foreclosed house : foreclosure : Foreign exchange : Foreign : Foreign investment : forex.com : Forex Future : Forex : Forex Option : Forex Swap : forward FX contract : France : France income tax : Free Cash Flow : Free market : Freedom : Frugal : Fun : Fund : Futures : FX Future : FX : FX Option : FX Swap : fxcm : Gain : Gains : Gamble : Garden : Gas : GBP : GBP Rate : gbp/usd : GDP : Germany : Germany income tax : Ghetto : Gift : Gold : Gold price : Gold prices : Golf : Government bond : Government buy : Government purchase : Great Britain : Great depression : Greece : Greed : Growth : Happy : Headlines : Health : Hedge : Hedging : Holiday : Holland : Home : Honda : Hong Kong dollar : Hong Kong : Hong Kong stock : House : Housing : HTC : Hyundai : Immigration : Import : Income : Income tax : India : Inexpensive : Inflation : Inflation rate : Initial public offering : Insurance : Interbank : Intercontinental exchange : Interest : Interest rate : Intervention : Invest : Investment : IPO : Ireland : Irish : ISO : Italy : Italy income tax : Japan forex : Japan : Japan income tax : Japanese Yen : Job market : JPY : KFC : Kitchen : Korea : Land : Leasing : Lending rate : Leverage : Libor : Lifestyle : Liquidation : Loan : London exchange market : London forex : London fx : London Libor : London stock exchange : London stock market : Long Position : Loss : Luxury : M&A : Make money : Margin : Market : Market Maker : Marriage : Mazda : Merger : Metals : Migration : Mining : Mitsubishi : MLB : Monetary : Monetary policy : money.cnn.com : Money CNN : Money : Mortgage : Movie : MSCI : NAFTA : Nasdaq : National Currency : NATO : NBA : Netherlands : New York Stock Exchange : Nissan : Notebook : NYSE : OANDA : Office : Oil : Oil price : OPEC : Open Position : Operating assets : Operating lease : Opportunity : OTC : OTC markets : Over the counter : Overbought : Oversold : Partnership : Petroleum : Pink sheet stocks : PIP : Pizza Hut : Plus 500 : Political party : Politics : Pollution : Poor : Portfolio : Portugal : Pound account : Pound Rate : Pound to Dollar : Poverty : Prediction : Premium : Present : President : Price : Price Movement : Price of gold : Prices : Privileged : Profit : Profitable : Real Estate : Real Exchange Rate : Recession : Relative Strength : Religion : Restaurant : Retail Prices : Return : Returns : Rich : Ring : Risk : Ruin : Russia : S&P500 : Samsung : Save : Savings : Schengen : Scotland : Scottish : SDR : Securities : Sell : Server : Share market : Shareholder : Shell : Shipping : Short Position : Short sell : Shortage : Sickness : Silicon Valley : Singing : Soccer : Social security card : Social security : Social security office : Society : Solar energy : SP500 : Space : Spain : Special Drawing Rights : Spend : Sports : Spot Forex : Spot FX : Sterling : Stock : Stock market : Stock symbol : Stocks : Stop buy : Stop loss : Success : Supplier : Swap : Swiss Franc : Table tennis : Tariffs : Tax benefit : Tax cut : Tax : Technology : Tennis : Terror : TMX : Tokio : Tourism : Toyota : Trade deficit : Trade : Trade union : Trader : Trading Model : Train : Transaction : Transaction Risk : Travel : Trip : UK bonds : UK Central Bank : UK GDP : UK government bonds : UK : UK income tax : UK inflation : UK interest rate : Unemployed : Unemployment : United Kingdom : United Kingdom income tax : United Kingdom tax : United States income tax : US bonds : US dollar : US GDP : US government bond : US income tax : US inflation : US interest rate : US

tax : USD : usd to gbp : Vacation : Valuable : Vendor : Volatility : Volleyball : Vote : War : Water : Wireless : Work : Working capital : World : World markets : Yahoo finance : Yen : Yield curve : Yield : Yuan

Note: Predictor variables consist of English-language Google Trends keywords. Authors made a list for relevant financial terms, abbreviations, stock market and foreign exchange market terminologies, along with the names of the famous financial news sites, and etc.

Figure 6: Illustration for How to Download Google Trends Search Volume Indices



Source: Google Trends

8. Appendix 3: Model Output

Table 7: Economic Classification of Model Inputs from Cluster Analysis, Rolling Origin

Cluster 1			**Cluster 2**		
Input Variables	**Frequency**	**Tag**	**Input Variables**	**Frequency**	**Tag**
Swiss_Franc_in_US	487	safehaven	EU_in_UK	462	trade
Pound_account_in_UK	385	output	Government_bond_in_UK	421	rate
Oil_price_in_US	377	price	Dollar_Rate_in_US	308	rate
Initial_public_offering_in_UK	373	rate	Initial_public_offering_in_UK	296	rate
Schengen_in_UK	293	trade	EUR_in_US	167	rate
Bank_of_England_in_UK	186	rate	Color_in_UK	120	output
Crude_Oil_in_US	126	price	GBP_in_US	116	rate
Charity_in_UK	123	output	Prediction_in_UK	84	rate
Color_in_UK	59	output	Crude_Oil_in_US	64	price
Money_in_US	40	risk	Schengen_in_UK	57	trade
gbp_slashsign_usd_in_UK	27	rate	Chevron_in_US	54	price
Mining_in_US	21	output	Charity_in_UK	39	output
Accountant_in_US...6	20	output	Over_the_counter_in_US	33	rate
England_in_US	19	trade	Recession_in_US	31	risk

EUR_in_US	19	rate	Bomb_in_UK	24	risk
Government_bond_in_UK	18	rate	Depression_in_UK	24	risk
JPY_in_UK	16	rate	Feder-al_income_tax_in_UK	22	rate
Chevron_in_US	15	price	Convert_currency_in_US	21	rate
GBP_in_US	13	rate	Headlines_in_UK	14	risk
Honda_in_US	13	trade	Trader_in_US	14	rate
Save_in_US	13	output	Liquidation_in_UK	13	risk
US_government_bond_in_UK	13	rate	fxcm_in_UK	12	rate
Trade_deficit_in_UK	12	rate	GBP_in_UK	12	out-put
Unit-ed_States_income_tax_in_UK	10	rate	EU_in_US	11	trade
Dollar_Rate_in_US	9	output	Home_in_UK	11	out-put
Shortage_in_UK	9	output	Import_in_US	11	trade
Train_in_UK	9	output	Bank_of_England_in_UK	10	rate
Euro_Area_in_US	8	rate	Daily_fx_in_UK	10	rate
Short_sell_in_US	8	risk	Lifestyle_in_US	10	out-put
Tourism_in_UK	8	trade	PIP_in_US	10	risk
Bankrupt_in_US	7	risk	Futures_in_US	9	rate
Election_in_UK	7	risk	Petroleum_in_US	9	price
Fixed_exchange_rate_in_UK	7	rate	Bear_market_in_UK	8	risk
Scottish_in_UK	7	output	Pound_to_Dollar_in_UK	8	rate
Shortage_in_US	7	output	Trade_deficit_in_UK	8	rate
GBP_in_UK	6	output	usd_to_gbp_in_US	8	rate
Ring_in_US	6	output	Arbitrage_in_UK	7	rate
Risk_in_UK	6	risk	IPO_in_UK	7	rate
Home_in_UK	5	output	Libor_in_US	7	rate
Audit_in_US	4	risk	Pound_account_in_UK	7	out-put
Bomb_in_UK	4	risk	NATO_in_UK	6	risk
Crash_in_UK	4	risk	Accountant_in_US...6	5	out-put
Daily_fx_in_UK	4	rate	Audit_in_US	5	risk
Energy_in_UK	4	price	Canadian_Dollar_in_UK	5	rate
Lon-don_stock_exchange_in_US	4	rate	Dollar_to_Pound_in_UK	5	rate
Mortgage_in_US	4	output	Margin_in_UK	5	rate
Prediction_in_UK	4	rate	NATO_in_US	5	risk
President_in_US	4	risk	OTC_in_UK	5	rate

Price_of_gold_in_UK	4	price	Stock_in_US	5	rate

Note: The most frequently selected Google keywords among the top 1,000 models in Cluster 1 and 2 of the rolling-origin design, identified through hierarchical clustering based on the Jaccard distance. Input variables are categorized/tagged into subgroups according to the authors' expert judgement.
Source: own estimation and classification

Table 8: Economic Classification of Model Inputs from Cluster Analysis, Rolling Window

Cluster 1					
Input Variables	**Frequency**	**Tag**	**Input Variables**	**Frequency**	**Tag**
EU_in_UK	911	trade	Schengen_in_UK	12	trade
Dollar_Rate_in_US	906	rate	Liquidation_in_UK	11	risk
Government_bond_in_UK	883	rate	Color_in_UK	10	output
Initial_public_offering_in_UK	832	rate	Dollars_in_UK	10	rate
GBP_in_US	578	rate	London_stock_exchange_in_UK	10	rate
Prediction_in_UK	127	rate	Holland_in_UK	9	trade
Chevron_in_US	58	price	Import_in_US	9	trade
usd_to_gbp_in_US	46	rate	Japan_income_tax_in_UK	9	rate
Bomb_in_UK	38	risk	Over_the_counter_in_US	9	rate
Foreign_exchange_in_UK	32	rate	PIP_in_US	9	rate
fxcm_in_UK	29	rate	Spain_in_UK	9	trade
Mazda_in_US	29	trade	Trader_in_US	9	rate
UK_government_bonds_in_UK	29	rate	Charity_in_UK	8	output
Stop_buy_in_US	28	risk	London_stock_exchange_in_US	8	rate
Futures_in_US	25	rate	Price_of_gold_in_UK	8	price
EU_in_US	24	trade	Privileged_in_UK	8	output
Shareholder_in_US	22	rate	British_stock_in_US	7	rate
US_dollar_in_UK	19	rate	Color_in_US	7	output
Transaction_in_US	18	output	Ford_in_UK	7	output
Volatility_in_US	14	risk	Forex_in_UK	7	rate
Libor_in_US	13	rate	Honda_in_US	7	trade
Operating_lease_in_UK	13	output	Lifestyle_in_US	7	output
Returns_in_US	13	rate	Short_sell_in_US	7	risk

Arbitrage_in_UK	12	rate	Bank_of_England_in_US	6	rate
Bourse_in_UK	12	rate	Bankrupt_in_US	6	risk

Note: The most frequently selected Google keywords among the top 1,000 models in Cluster 1 of the rolling-window design, identified through hierarchical clustering based on the Jaccard distance. Input variables are categorized/tagged into subgroups according to the authors' expert judgement.
Source: own estimation and classification

9. References

1. J.F. Bilson, "Rational Expectations and the Exchange Rate", The Economics of Exchange Rates (Collected Works of Harry Johnson), Routledge, 2013.
2. L. Bulut, "Google Trends and the Forecasting Performance of Exchange Rate Models", Journal of Forecasting, 2018, 37, 303–315.
3. M. Masuda, F. Takeda, "Application of Google Trends Data in Exchange Rate Prediction", International Symposium on Forecasting, 2019. https://isf.forecasters.org/wp-content/uploads/gravity_forms/2-dd30f7ae09136fa695c552259bdb3f99/2019/06/Exchange-Rate.pdf
4. J.A. Frenkel, "A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence", The Scandinavian Journal of Economics, 1976, 78, 200–224.
5. R. Dornbusch, "Expectations and Exchange Rate Dynamics", Journal of Political Economy, 1976, 84 (6), 1161–1176.
6. J.F.O. Bilson, "The Monetary Approach to the Exchange Rate: Some Empirical Evidence", Staff Papers – International Monetary Fund, 1978, 25 (1), 48–75.
7. J.A. Frankel, "On the Mark: A Theory of Floating Exchange Rates Based on Real Interest Differentials", American Economic Review, 1979, 69 (4), 610–662.
8. R.A. Meese, K. Rogoff, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?", Journal of International Economics, 1983, 14, 3–24.
9. M. Chojnowski, P. Dybka, "Is Exchange Rate Moody? Forecasting Exchange Rate with Google Trends Data", Econometric Research in Finance, 2017, 2 (1), 1–21.
10. H. Yaganti, Y. Manpuria, "Exchange Rate Forecast Enhancement Using Sentiments from Google Trends", The IUP Journal of Financial Risk Management, 2018, 15 (2), 7.
11. H. Choi, H. Varian, "Predicting the Present with Google Trends", The Economic Record, June 2012, 88, 2–9.
12. N. McLaren, R. Shanbhogue, "Using Internet Search Data as Economic Indicators", Bank of England Quarterly Bulletin, 2011, 51 (2), 134–140.
13. G. Tkacz, "Predicting Recessions in Real-Time: Mining Google Trends and Electronic Payments Data for Clues", CD Howe Institute Commentary, 2013, 387, 1–28.
14. A. Petropoulos, V. Siakoulis, E. Stavroulakis, P. Lazaris, N. Vlachogiannakis, "Employing Google Trends and Deep Learning in Forecasting Financial Market Turbulence", Journal of Behavioral Finance, 2022, 23 (3), 353–365.
15. T. Preis, H.S. Moat, H.E. Stanley, "Quantifying Trading Behavior in Financial Markets Using Google Trends", Scientific Reports, 2013, 3 (1684), 1–6.
16. Z. Da, J. Engelberg, P. Gao, "The Sum of All FEARS Investor Sentiment and Asset Prices", The Review of Financial Studies, 2015, 28 (1), 1–32.

17. H. Hu, L. Tang, S. Zhang, H. Wang, "Predicting the Direction of Stock Markets Using Optimized Neural Networks with Google Trends", Neurocomputing, 2018, 285, 188–195.
18. M.H. Fan, M.Y. Chen, E.C. Liao, "A Deep Learning Approach for Financial Market Prediction: Utilization of Google Trends and Keywords", Granular Computing, 2021, 6 (1), 207–216.
19. A. Basistha, A. Kurov, M.H. Wolfe, "Volatility Forecasting: The Role of Internet Search Activity and Implied Volatility", Journal of Risk Model Validation, 2020, 14 (1), 35–63.
20. A.A. Salisu, A.E. Ogbonna, I. Aded, "Stock-Induced Google Trends and the Predictability of Sectoral Stock Returns", Journal of Forecasting, 2021, 40 (2), 327–345.
21. M.D. Bustamante, A.H.D. Valle, A.O. Ramirez, "The Google Trends Effect on the Behavior of the Exchange Rate Mexican Peso-US Dollar", Contaduría y Administración, 2019, 64 (2), 1–14.
22. K. Cortez, "Predicting Foreign Exchange in Emerging Markets with a Nearest Neighbor Approach: Fundamentals versus Online Attention Indicators", Financial Innovation, 2025, 11 (125), 1–20.
23. M. Evans, R. Lyons, "Order Flow and Exchange Rate Dynamics", Journal of Political Economy, 2002, 110 (1), 170–180.
24. M. Evans, R. Lyons, "How Is Macro News Transmitted to Exchange Rates?", Journal of Financial Economics, 2008, 88 (1), 26–50.
25. M. Dzielinski, "Measuring Economic Uncertainty and Its Impact on the Stock Market", Finance Research Letters, 2012, 9 (3), 167–175.
26. J. Kogan, C. Nicholas, M. Teboulle, Grouping Multidimensional Data, Berlin Heidelberg New York: Springer, 2006.
27. S.S. Choi, S.H. Cha, C.C. Tappert, "A Survey of Binary Similarity and Distance Measures", Journal of Systemics, Cybernetics and Informatics, 2010, 8 (1), 43–48.
28. F. Batool, C. Hennig, "Clustering with the Average Silhouette Width", Computational Statistics & Data Analysis, 2021, 158, 107190.
29. G.P. Smith, "Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency", Finance Research Letters, 2012, 9 (2), 103–110.
30. M. Wilson, A. Yelowitz, "Characteristics of Bitcoin Users: An Analysis of Google Search Data", MPRA Paper, 2014, 59661, 1–13.
31. G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, 8th ed., New York: Springer, 2017.
32. M. Verbeek, A Guide to Modern Econometrics, 4th ed., Wiley, 2012.
33. F.S. Mishkin, "The Foreign Exchange Market", in The Economics of Money, Banking, and Financial Markets, 7th ed., Boston: Pearson Addison Wesley, 2004, 435–461.
34. W. Enders, Applied Econometric Time Series, 4th ed., Wiley, 2015.
35. J.F.O. Bilson, "The Deutsche Mark/Dollar Rate: A Monetary Analysis", Carnegie-Rochester Conference Series on Public Policy, 1979, 11, 59–101.
36. R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed., Melbourne, Australia: OTexts, 2018.