NOVA IMS
Information Management School

MEGI

Master Degree Program in
Statistics and Information Management

Multi-country Analysis of Unemployment Rate Nowcasting During Covid-19 With Search Query Data

Arthur Henrique Fernandes Campos

Dissertation

presented as partial requirement for obtaining the Master Degree Program in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

MULTI-COUNTRY ANALYSIS OF UNEMPLOYMENT RATE NOWCASTING DURING COVID-19 WITH SEARCH QUERY DATA

By

Arthur Henrique Fernandes Campos

Master Thesis presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Risk Analysis and Management.

Co-Supervisor*:* Bruno Damásio

Co-Supervisor*:* Flávio Pinheiro

February 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

São Paulo, 27 de fevereiro de 2023.

ABSTRACT

Nowcasting methods aim to predict the present and the very near future and past to circumvent data lag. As internet usage becomes ubiquitous, more and more individuals use internet search engines as decision-making tools; consequently, search query data may be good proxies for individual behavior, and thus a useful nowcasting predictor variable for many macroeconomic indicators. This study examines the potential of using Google Trends data to nowcast unemployment rate during the years of the Covid-19 pandemic across sixteen countries by comparing the performance of four alternative models with Google Trends data against a base autoregressive model, considering two modelling training windows, one limited to pre-Covid data and the other including 2020 data. The results show that search query data lack robustness and have varying predictive power, with the inclusion of 2020 data into the training set providing a significant improvement of out-of-sample forecasting accuracy. These findings indicate that search query data may have good predictive power in some scenarios, but may not be robust enough for real-life applications.

KEYWORDS

Economic forecasting; Nowcasting; Unemployment; Google Trends.

INDEX

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

From personal budgeting and business planning to government policymaking, throughout many facets of modern society, individuals and institutions alike rely on official statistics for decision-making. However, given the work involved in data collection, treatment, and analysis, all official releases are bound to be released with a delay (i.e., results for February may only be available in March). When accurate present information is imperative, it proves necessary to forecast the present—the “now”.

Nowcasting aims to forecast not the medium- or long-term future, but what is happening right now. Often, statisticians include more recent data into a framework; for instance, to predict a monthly indicator, it may be useful to include results from a weekly survey, or even a covariate (i.e., a predictor) with the same release frequency but different release window. Overall, the motto is: timely information is strategic.

With the widespread adoption of the Internet and the technological advancements that followed the digital age, a vast amount of public data became readily available online. One such data source, the search query aggregator Google Trends, allows users to track the interest in specific topics and categories across geography and time. Under the hypothesis that search engine queries are a reasonable proxy for actual behaviors (e.g., purchasing habits), monitoring changes in these queries might give insight into current events and conditions that might otherwise take weeks or months to be reflected in official statistics. Thus, should said hypothesis prove correct, then the potential exists in using Google Trends (GT) data to refine nowcasting results and make better decisions faster.

Since the seminal works on the topic by Choi and Varian (2009, 2012), many researchers have explored the idea further, to varying degrees of success. One constant, however, has been the prevalence of studies focused on the developed world, even though developing countries often suffer from longer release lags or less reliable covariate data altogether.

This study focus on exploring the potential of GT data for nowcasting unemployment rate in 16 countries including high- and low-income economies. These are Australia, Brazil, Canada, Chile, Germany, Italy, Japan, Mexico, the Netherlands, Portugal, South Korea, Switzerland, Turkey, the United Kingdom, the United States, and Uruguay. These countries have easy-to-access monthly unemployment data releases from primary sources and, apart from Uruguay, are among the 50 largest economies by nominal GDP.

The dissertation is organized as follows. Chapter 2 contains an overview of current literature on the topics of nowcasting and forecasting with search query data. Chapter 3 describes the data and its collection. Chapter 4 goes into the nowcasting methodology. Chapter 5 presents and discusses the nowcasting results per scenario. Finally, Chapter 6 concludes the work with a brief reflection upon the results and suggestions for further exploration of the theme.

2. LITERATURE REVIEW

While back at its launch Google Trends was perceived as a tool for webmasters and marketers alike for search engine optimization purposes, starting in 2008 Google Inc. and independent authors published papers on applying search engine aggregated data for scientific research in epidemiology (Polgreen et al. 2008) and economic nowcasting (Choi & Varian, 2009).

In September 2008, Google Inc. released Google Flu Trends (GFT), a flu nowcasting service fueled by Google Trends queries. Until its shut down on August 09, 2015, GFT was the topic of many discussions regarding its predictive power and utility as an outbreak prediction tool. Notably, Olson et al. (2013) found that, even with the 2009 revised methodology, GFT fell short by 52% in its influenza-like infections prediction for New York City during the 2009 A/H1N1 pandemic. But, in a recent turn of events, Kandula and Shaman (2019) look at new surveillance data to reevaluate the GFT estimation errors and create a random forest regression model with GFT rates that see an error reduction of 80% to the 2012/13 season original prediction, suggesting a reevaluation of search query usage as predictor variables in influenza forecast systems.

In economics and business, however, reception of GT data-powered forecasts and nowcasts has been more favorable, seeing use for instance in predicting tourism inflows (Artola, Pinto & de Pedraza García, 2015) and demand (Siliverstovs & Wochner, 2018), suicide occurrences (Kristoufek, Moat & Preis, 2016), and fashion consumer behavior for a big player in the industry (Silva et al., 2019). In the topic of financial markets, GT data have been used for predicting downturn stock market moves (Preis, Moat & Stanley, 2013), foreign exchange rates (Bulut, 2017), direction of opening stock prices (Hu et al., 2018) and accruable returns on precious metals (Salisu, Ogbonna & Adewuyi, 2020).

Some authors are less enthusiastic about the prospect of Google Trends as a predictor. Nagao, Takeda and Tanaka (2019) suggest that GT data-driven models may lack robustness and are dependent on data frequency and seasonality adjustments with no consistency regarding whether they would improve or reduce accuracy. Schaer, Kourentzes and Fildes (2019) find that established forecasting benchmarks outperform those with GT data and social network information in forecasting video game sales and corporate online video views, although the authors acknowledge they are limiting the analysis to linear models.

When it comes to a focus on macroeconomic variables, there seems to be a natural tendency to study dependent variables closely related to individual behavior, namely private consumption, and unemployment; Vosen and Schmidt (2011), Choi and Varian (2012), Vosen and Schimidt (2012), Carrière-Swallow and Labbé (2013), and Woo and Owen (2019) look at the former, while Choi and Varian (2009), Barreira, Godinho and Melo (2013), Fondeur and Karamé (2013), Vicente, López-Menéndez and Pérez (2015) and Naccarato et al. (2018) at the latter. As GT data pertains mostly to individual search queries, it is expected that forecasting models for those variables would benefit from such data. Fewer studies, such as Marcellino and Schumacher (2010), Kuzin, Marcellino and Schumacher (2011), and Bantis, Clements and Urquhat (2021), look at GDP growth rate.

As for location, the economies studied, for the most part, are part of the geopolitical so-called developed world; the exceptions being Brazil (Bantis, Clements & Urquhat, 2021), Chile (Carrière-Swallow & Labbé, 2013) and Hong Kong, China (Choi & Varian, 2012). Carrière-Swallow and Labbé (2013) point out that good nowcasting methods are even more important in developing countries as the lag before data release is often longer in those countries than in the developed world. Interestingly though, of those three only Brazil is not classified as a high-income economy by the World Bank, meaning there is plenty of room for innovation on the topic.

Google Trends category and keyword selection varies widely between works. In the topic of private consumption, most authors select multiple categories related to different goods, most often tying together with the types of goods regarded by the index to be nowcasted. The biggest innovation comes from Woo and Owen (2019) by splitting apart categories related to durable and non-durable goods and services consumption, and adding keywords related to economic recessions, limited to news only, as a reflection of the relationship between consumption and news media. For unemployment rate, using keywords rather than broad categories are the norm, with some authors exploring a simple, single query such as ‘jobs’ (D’Amuri & Marcucci, 2017) or ‘job offer’ (Naccarato et al., 2018); on the other hand, Mulero and García-Hiernaux (2021) use queries related to leading job search applications and websites, national unemployment centers, usual job searching terms (e.g., ‘how to find a job’) and companies with large headcount to forecast Spanish unemployment. Meanwhile, to forecast GDP growth rates in Brazil and the United States, Bantis, Clements and Urquhat (2021) simply look at all main categories and first level subcategories available on the platform.

When it comes to testing methodology, most studies boil down to a comparison between a baseline (vanilla) autoregressive model and an alternative model including GT data. Carrière-Swallow and Labbé (2013) expand the baseline models, looking at AR(1), AR(1) including a commonly used proxy variable, and the best performing ARMA(p,q) model on the basis of Akaike information criterion (AIC) and the Bayesian information criterion (BIC), depending on the sample size, and comparing them to alternative models including GT data. Meanwhile, Choi and Varian (2009, 2012) instead use multiple alternative models including different categories to find the best fitting predictor set. Vosen and Schmidt (2011) go further and perform the same analysis to other competing proxy variables as well to see not only whether search query data has predictive power, but also how it compares against usual benchmarks. Straying from the formula, Fondeur and Karamé (2013) apply diffuse Kalman filter models, such that stochastic trends are represented with a random walk with a time-varying drift; the baseline model estimates both the studied time series and the Google Trends series simultaneously but independent of each other, while the alternative, bivariate model is such that the studied time series slope instantaneously depends on the Google slope.

There are alternative methodological approaches to modelling, however. Ghysels et al. (2016) present the R statistical package midasr for mixed-data sampling (MIDAS) regression models with independent variables of different sampling frequencies. Even though mixed-frequency vector autoregression (MF-VAR) models also account for multiple sampling frequencies, Kuzin, Marcellino and Schumacher (2011) compare both approaches with euro GDP nowcasts and conclude MIDAS models show better performance for shorter time horizons. Such mixed-frequency models may be a good fit for Google Trends-driven models should the researcher adopt weekly data frequency to predict monthly indicators, or monthly data frequency to predict quarterly or yearly indicators. Another modelling option is a multivariate Markov chain (MMC), which may be used to improve forecasts in the scenario of a predictor variable with unknown values in the forecasting period (Damásio & Nicolau, 2014) and together with VAR models to better capture non-linearity (Damásio & Mendonça, 2019, 2022). One advantage of MMC modelling is the detection of structural breaks that could otherwise negatively impact the forecasts; Damásio & Nicolau (2020) present a method to detect multiple such breaks.

Some studies branch out from Google Trends and include other online data sources. Elshendy et al. (2018) analyze the relationship between crude oil price and multiple predictors from Google Trends, Twitter, Wikipedia and the Global Data on Events, Location and Tone (GDELT) database, finding evidence to the advantage of integrating multiple different platforms in the predictive models rather than just one. Weng et al. (2018) apply sentiment analysis to published news articles, Google search queries and Wikipedia unique visitors data on top of traditional sources to train an AI platform for predicting 1-day ahead stock prices, concluding that online data is significant but not consistent to prediction accuracy.

Finally, on the topic of proper study documentation, Nuti et al. (2014) compile health care research using GT data and find that only 7% of the selected studies include enough search query information to be reproducible. To tackle this problem going forward, the authors propose a simple but sufficient checklist to insure study reproducibility.

3. DATA

Google Trends data are time series of the interest over time of a query. This query can be a literal string (e.g., the string “job opportunities” will limit results to “job opportunities”), a topic (e.g., the topic “World Cup” will encompass results related to the football competition colloquially known as “World Cup”), or a category (e.g., the category “Candy & Sweets” will consider results related to candy and sweets). The latter can be combined with the first two to further filter the queries. For topics and categories, the values returned are based on query categorization according to the Google Inc. internal natural language processing model.

It is possible to limit the search to a specific time window, to a specific country or country subdivision, or to specific Google search engines (e.g., image search). It is also possible to compare multiple queries at the same time.

The time series values are expressed as integer values ranging from 0 to 100, in which 100 is equivalent to the maximum interest observed in the filtered timeframe, and all other values are equal to the percentage of interest compared to the maximum; that is, a value of X means that the relevant query was searched about X% as often as when it was most searched in the selected period.[1] Therefore, results from disjoint timeframes or separate results from different queries are incompatible with one another.

To serve this information in a timely manner, instead of returning the true interest over time of a query (i.e., the real volume of a query versus all other queries), Google Trends takes a sample of all queries and then returns the sampled interest over time (i.e., the sampled volume of a query versus all sampled queries). A consequence of this pre-processing is that all values have some sampling error. As Medeiros and Pires (2021) show, the results of any analysis with Google Trends data may vary depending on the day of collection and lead to different results than the real value would provide.

The simplest way to address this issue is to retrieve Google Trends data with the same parameters from multiple different samples and then to take the mean of all observations of the same period. The problem, then, lies on getting results from many different samples. While there is no API for the Google Trends platform, there are many unofficial packages for popular programming languages available that allow users to automate data collection.

It used to be the case that all data requests in a period of 24 hours from neighboring IP addresses would use the same sample and thus have the same data points, but this behavior changed on February 16th, 2022. Since then, even consecutive data requests can differ in at least one data point; the longer the delay between requests, the closer the difference between requests mimics what was seen before this change. Further details on past and current behavior are found in Appendix A.

For the study, the final Google Trends time series for each category and country is the arithmetic mean of 79 series, pruned from a total of 102 series collected with the gtrendsR package between August 2022 and October 2022. The pruning aims to discard series too similar to others due to the new sampling behavior and is described in Appendix A. The data requests looked at interest over time from

[1] When the maximum interest is above a certain threshold compared to the minimum interest, some data points are presented as “<1”.

2004 to 2021 of categories 60 and 706, respectively "Jobs & Education/Jobs" and "Law & Government/Social Services/Welfare & Unemployment", with all search engines enabled and no query selected as to capture as much relevant data as possible, for each selected country. This category selection follows the work of Choi and Varian (2009).

As for dependent variables, all unemployment rate data used is from national primary sources, apart from thirteen data points (Turkey, Dec/12–Dec/13) extracted from Eurostat for completeness. Table 3.1 lists all sources used per country. Brazil and the United Kingdom publish unemployment rate as three-month aggregates; this study sets the aggregate as the value of the third month (e.g., May/19–Jul/19 is set to Jul/19). As it may be impossible to reconcile the data regarding any methodological treatment performed before release, none of the series go through further adjustments. The final data are monthly, ranging from December 2012 to December 2021, divided into unadjusted and seasonally adjusted time series according to the series available for each country. In total, the study encompasses 12 countries with unadjusted time series and 12 countries with seasonally adjusted time series, for a final count of 16 countries, of which three are not classified as high-income economies by the World Bank: Brazil, Mexico, and Turkey.

Table 3.1 – Data sources and type of time series available for unemployment rate, per country

Country	**Data source**	**Time series available**
Australia	Australian Bureau of Statistics, Labour Force Survey	Unadjusted and seasonally adjusted
Brazil	Instituto Brasileiro de Geografia e Estatística (IBGE), Pesquisa Nacional por Amostra de Domicílios Contínua	Unadjusted
Canada	Statistics Canada. Table 14-10-0017-01. Labour force characteristics by sex and detailed age group	Unadjusted
Chile	Instituto Nacional de Estadísticas (INE), Encuesta Nacional de Empleo (ENE)	Unadjusted and seasonally adjusted
Germany	Bundesagentur für Arbeit	Unadjusted
Italy	Istituto Nazionale di Statistica (ISTAT), Labour and wages	Seasonally adjusted
Japan	Statistics Bureau of Japan, Labour Force Survey	Unadjusted and seasonally adjusted
Mexico	Instituto Nacional de Estadística y Geografía (INEGI), Encuesta Nacional de Ocupación y Empleo (ENOE)	Seasonally adjusted
Netherlands	Statistics Netherlands, Dutch Labour Force Survey, Monthly labour participation and unemployment	Unadjusted and seasonally adjusted

Country	Data source	Time series available
Portugal	Statistics Portugal, Labour force survey	Unadjusted and seasonally adjusted
South Korea	Statistics Korea, Economically Active Population Survey	Unadjusted and seasonally adjusted
Switzerland	Swiss Federal Statistical Office (SFSO), Swiss Labour Force Survey	Seasonally adjusted
Turkey	Eurostat (data from Dec/12–Dec/13) Turkish Statistical Institute (TURKSTAT), Labour Force Statistics (data from 2014–2021)	Seasonally adjusted
United Kingdom	Office for National Statistics, Labour Force Survey	Unadjusted and seasonally adjusted
United States	United States Department of Labor, Labor Force Statistics from the Current Population Survey (CPS)	Unadjusted and seasonally adjusted
Uruguay	Instituto Nacional de Estadística (INE), Encuesta Continua de Hogares	Unadjusted

All unemployment rate values are in percentages, in a range of 0 to 100, and preserve any rounding already present in the original data.

As a last step, all unemployment rate and Google Trends time series go through the Kwiatkowski–Phillips–Schmidt–Shin test to determine whether any differentiation is required for trend-stationarity. The test results show that lag-1 and lag-12 differentiations are sufficient to ensure all series are trend-stationary; regardless of any previous seasonal adjustment, all series are differentiated as such. Mathematically, this is expressed as

$$\Delta y_t = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13}) \tag{4.1}$$

$$\Delta GT_{x,t} = \left(GT_{x,t} - GT_{x,t-1}\right) - \left(GT_{x,t-12} - GT_{x,t-13}\right) \tag{4.2}$$

in which y_t is the original unemployment rate series for period t, $GT_{x,t}$ is the original Google Trends series for category x and period t, and Δy_t and $\Delta GT_{x,t}$ are the differentiated unemployment rate and Google Trends series respectively.

4. METHODOLOGY

Due to the often-seen abrupt rise in unemployment following the beginning of the Covid-19 pandemic, it is possible to test the predictive power of search query data in a stress scenario while still using contemporary data from a time frame when Internet access has become more widespread. This can be done by limiting the training set to pre-Covid data and then running out-of-sample forecasts for the Covid-19 period. One other scenario of interest can be achieved by including Covid-19 data in the training set to test if the predictive power of the independent variables increases or decreases. These two scenarios may provide insight into search query data as predictor variables and their robustness. Thus, this study considers two training windows, each comprising of sixty observations, ending in the Decembers of 2019 and 2020.

Taking inspiration in parts by the methodological approach outlined by Vosen and Schmidt (2011), a base model with past values of the dependent variable works as the starting point and sanity check. This model is then expanded into multiple alternative models that include Google Trends data.

Thus, for each country, variable and training window, there are five competing models: (1) an AR(1) component without GT data, (2) linear regression on GT data, (3) linear regression on GT data including lagged categories, (4) linear regression on GT data and an AR(1) component, and (5) linear regression on GT data including lagged categories and an AR(1) component. This framework is an expansion upon the work of Choi and Varian (2009). Mathematically, the models may be expressed, in order, as

$$\Delta y_t = \alpha_1 + \rho_1 \Delta y_{t-1} + \lambda_t \tag{5.1}$$

$$\Delta y_t = \alpha_2 + \delta_{2,0} \Delta GT_{60,t} + \psi_{2,0} \Delta GT_{706,t} + \tau_t \tag{5.2}$$

$$\Delta y_t = \alpha_3 + \delta_{3,0} \Delta GT_{60,t} + \psi_{3,0} \Delta GT_{706,t} + \delta_{3,1} \Delta GT_{60,t-1} + \psi_{3,1} \Delta GT_{706,t-1} + \nu_t \tag{5.3}$$

$$\Delta y_t = \alpha_4 + \rho_4 \Delta y_{t-1} + \delta_{4,0} \Delta GT_{60,t} + \psi_{4,0} \Delta GT_{706,t} + \varphi_t \tag{5.4}$$

$$\Delta y_t = \alpha_5 + \rho_5 \Delta y_{t-1} + \delta_{5,0} \Delta GT_{60,t} + \psi_{5,0} \Delta GT_{706,t} + \delta_{5,1} \Delta GT_{60,t-1} + \psi_{5,1} \Delta GT_{706,t-1} + \epsilon_t \tag{5.5}$$

in which Δy_t is the differentiated unemployment rate series for period t; α_i, ρ_i, $\delta_{i,p}$, and $\psi_{i,p}$, are model parameters for model i and lag p; λ_t, τ_t, ν_t, φ_t, and ϵ_t are white noise series for period t; and $\Delta GT_{x,t}$ is the differentiated Google Trends series for category x and period t. The preferred model is then chosen based on whichever achieves the lowest AIC value.

Finally, out-of-sample forecasts are set up as one-step forecasts (i.e., no previous estimations are used to calculate new ones and instead use actual values). This emulates a real-world nowcasting application, in which only short-term forecasts are of interest and all most recent data is available to empower such forecasts. The measure of error adopted for comparing forecasts is the root mean squared error (RMSE).

5. RESULTS AND DISCUSSION

5.1. SCENARIO A: 2015–2019

This scenario separates pre-Covid and Covid-19 data into training and testing sets respectively. Should the alternative models be robust, then they should produce more accurate 2020 and 2021 forecasts than the base model.

Five out of 12 countries with unadjusted time series and six out of 12 countries with seasonally adjusted time series had alternative models with lower AIC than the base model, for a total of seven countries: Italy, Japan, Portugal, South Korea, the United Kingdom, the United States, and Uruguay. Table 5.1 presents the AIC values for each model.

Table 5.1 – AIC values per model per variable per country, Scenario A

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	-196.56	-190.26	-191.17	-192.77	-193.96
	SA	-212.16	-208.35	-209.86	-208.76	-210.35
Brazil	U	-223.78	-185.65	-181.83	-220.90	-217.50
Canada	U	-175.88	-172.53	-168.55	-172.79	-168.84
Chile	U	-178.74	-178.07	-175.41	-176.11	-173.48
	SA	-178.97	-178.34	-175.85	-176.43	-173.97
Germany	U	-328.09	-322.74	-318.86	-325.59	-321.66
Italy	SA	-112.84	-99.29	-101.13	-114.48	**-116.60**
Japan	U	-211.99	-201.84	-199.11	**-212.59**	-209.27
	SA	-212.64	-209.97	-208.47	**-216.23**	-212.65
Mexico	SA	-168.22	-148.83	-144.94	-164.38	-160.62
Netherlands	U	-232.18	-228.24	-226.55	-228.70	-226.78
	SA	-235.63	-233.99	-232.99	-233.53	-232.91
Portugal	U	-175.49	**-175.65**	-172.03	-173.99	-170.35
	SA	-173.91	**-176.00**	-173.43	-174.58	-171.78
South Korea	U	-145.08	-138.84	-137.06	**-148.73**	-145.23
	SA	-149.41	-147.39	-145.69	**-153.10**	-149.67
Switzerland	SA	-211.66	-190.85	-187.97	-208.51	-206.92
Turkey	SA	-116.72	-97.09	-95.54	-113.22	-110.50
United Kingdom	U	-279.04	-276.83	-278.94	-275.13	-277.84
	SA	-280.08	-277.59	**-280.09**	-276.12	-279.33
United States	U	-210.75	-205.17	-207.09	-211.44	**-211.54**
	SA	-190.77	-191.32	**-195.56**	-192.48	-195.01
Uruguay	U	-17.71	-13.36	-11.72	**-22.68**	-20.77

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
Bold and underscored values indicate alternative models that are a preferred model.

At least one alternative model beats the base model in out-of-sample performance in eight countries for 2020 data and in nine countries for 2021 data: Australia (2020, 2021), Canada (2020, 2021), Chile (2020, 2021), Germany (2020, 2021), Japan (2020), Mexico (2020, 2021), the Netherlands (2020, 2021), Turkey (2020, 2021), South Korea (2021), and Uruguay (2021).

When limiting the data to preferred models, there are two cases in which the preferred model outperforms the base model in out-of-sample performance: the seasonally adjusted series for Japan in 2020 and the unadjusted series for Uruguay in 2021, with respective changes in RMSE of -1,6% and -3,9%.

On the other end of the scale, the preferred model in the seasonally adjusted series for Italy produces very poor forecasts compared to the base, with increases in RMSE of 193,2% in 2020 and 154,4% in 2021. Meanwhile, another alternative model with lower AIC than the base fares better than the preferred model, reducing the difference in RMSE versus the base model to 11,7% and 2,5% in 2020 and 2021 respectively.

Excluding Italy as an outlier, in 2020 data the percentage difference in RMSE goes from -1,6% to 27,5%, while in 2021 data it ranges from -3,9% to 24,1%. Table 5.2 and Table 5.3 present the out-of-sample forecast RMSE values for 2020 and 2021 respectively.

Table 5.2 – Out-of-sample forecast RMSE values and percentage differences for 2020, Scenario A

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	0.46729	0.41738	0.44490	0.46807	0.50334
			-10.68%	-4.83%	0.17%	7.72%
	SA	0.46722	0.43739	0.46905	0.47175	0.50447
			-6.38%	0.39%	0.97%	7.97%
Brazil	U	0.30873	0.64435	0.70025	0.37633	0.38997
			108.71%	126.82%	21.90%	26.31%
Canada	U	2.11369	1.91580	1.89169	2.12371	2.13443
			-9.36%	-10.50%	0.47%	0.98%
Chile	U	0.82059	0.81744	0.77748	0.83507	0.80104
			-0.38%	-5.25%	1.76%	-2.38%
	SA	0.78695	0.77988	0.74368	0.80371	0.77472
			-0.90%	-5.50%	2.13%	-1.55%
Germany	U	0.31225	0.28212	0.28065	0.31491	0.31277
			-9.65%	-10.12%	0.85%	0.17%
Italy	SA	0.94008	1.82280	3.15852	1.05023	**0.75601**
			93.90%	235.98%	11.72%	**193.17%**
Japan	U	0.14343	0.13911	0.14172	**0.14585**	0.13323
			-3.01%	-1.19%	**1.68%**	-7.11%
	SA	0.16064	0.14468	0.16700	**0.15801**	0.16106
			-9.94%	3.96%	**-1.64%**	0.26%
Mexico	SA	0.52435	0.64092	0.63790	0.52890	0.52415

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
			22.23%	21.65%	0.87%	-0.04%
Netherlands	U	0.31453	0.278252	0.279894	0.31076	0.31044
			-11.53%	-11.01%	-1.20%	-1.30%
	SA	0.29893	0.27186	0.28046	0.28904	0.29979
			-9.06%	-6.18%	-3.31%	0.29%
Portugal	U	0.56934	**<u>0.60007</u>**	0.61227	0.59183	0.59781
			<u>5.40%</u>	7.54%	3.95%	5.00%
	SA	0.55292	**<u>0.60245</u>**	0.63681	0.59573	0.62312
			<u>8.96%</u>	15.17%	7.74%	12.69%
South Korea	U	0.34768	0.49142	0.52016	**<u>0.43853</u>**	0.45705
			41.34%	49.61%	**<u>26.13%</u>**	31.46%
	SA	0.28848	0.40153	0.42750	**<u>0.36783</u>**	0.38137
			39.19%	48.19%	**<u>27.50%</u>**	32.20%
Switzerland	SA	0.13944	0.21358	0.24823	0.18268	0.23603
			53.17%	78.03%	31.01%	69.27%
Turkey	SA	0.76704	0.74723	0.88406	0.86881	0.97673
			-2.58%	15.26%	13.27%	27.34%
United Kingdom	U	0.17210	0.18153	0.17881	0.17353	0.17261
			5.48%	3.90%	0.83%	0.30%
	SA	0.17141	0.17992	**<u>0.17615</u>**	0.17162	0.17233
			4.96%	**<u>2.77%</u>**	0.12%	0.54%
United States	U	3.44635	4.12511	3.96150	4.04447	**<u>4.07108</u>**
			19.69%	14.95%	17.36%	**<u>18.13%</u>**
	SA	3.33147	4.19825	**<u>4.07052</u>**	4.04691	4.05539
			26.02%	**<u>22.18%</u>**	21.48%	21.73%
Uruguay	U	0.74317	0.92234	1.01664	**<u>0.77248</u>**	0.82870
			24.11%	36.80%	**<u>3.95%</u>**	11.51%

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
Percentage difference values are in relation to the base model.
Bold and underscored values indicate alternative models that are a preferred model.

Table 5.3 – Out-of-sample forecast RMSE values and percentage differences for 2021, Scenario A

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	0.70517	0.60934	0.66429	0.71141	0.77040
			-13.59%	-5.80%	0.88%	9.25%
	SA	0.70060	0.62743	0.68573	0.69659	0.75177
			-10.44%	-2.12%	-0.57%	7.30%
Brazil	U	0.32060	0.84588	0.90028	0.44096	0.45682
			163.84%	180.81%	37.54%	42.49%
Canada	U	2.01147	1.79371	1.76694	2.02401	2.03338
			-10.83%	-12.16%	0.62%	1.09%
Chile	U	0.99230	0.97980	0.97706	1.00165	1.00449

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
			-1.26%	-1.54%	0.94%	1.23%
	SA	0.96462	0.94404	0.93942	0.97539	0.97671
			-2.13%	-2.61%	1.12%	1.25%
Germany	U	0.35236	0.31285	0.31122	0.35408	0.35163
			-11.21%	-11.68%	0.49%	-0.21%
Italy	SA	1.04137	1.70160	2.99813	1.06516	**<u>2.64925</u>**
			63.40%	187.90%	2.28%	**<u>154.40%</u>**
Japan	U	0.19812	0.21534	0.22989	**<u>0.20430</u>**	0.21398
			8.69%	16.04%	**<u>3.12%</u>**	8.01%
	SA	0.17263	0.18672	0.19648	**<u>0.17883</u>**	0.18322
			8.16%	13.82%	**<u>3.59%</u>**	6.13%
Mexico	SA	0.56578	0.54537	0.54495	0.56812	0.57234
			-3.61%	-3.68%	0.41%	1.16%
Netherlands	U	0.37383	0.36204	0.32729	0.39018	0.36053
			-3.15%	-12.45%	4.37%	-3.56%
	SA	0.37494	0.41206	0.39819	0.42284	0.41850
			9.90%	6.20%	12.78%	11.62%
Portugal	U	0.56963	**<u>0.61858</u>**	0.63844	0.60836	0.61886
			<u>8.59%</u>	12.08%	6.80%	8.64%
	SA	0.59537	**<u>0.66631</u>**	0.71772	0.65725	0.69840
			<u>11.92%</u>	20.55%	10.39%	17.31%
South Korea	U	0.52175	0.47096	0.47508	**<u>0.52715</u>**	0.53720
			-9.74%	-8.95%	**<u>1.03%</u>**	2.96%
	SA	0.44550	0.41665	0.43138	**<u>0.44863</u>**	0.45105
			-6.48%	-3.17%	**<u>0.70%</u>**	1.24%
Switzerland	SA	0.15071	0.22922	0.26545	0.18945	0.23728
			52.09%	76.13%	25.70%	57.44%
Turkey	SA	1.04308	0.90335	1.17248	1.05918	1.28279
			-13.40%	12.41%	1.54%	22.98%
United Kingdom	U	0.25535	0.27667	0.28343	0.25747	0.26028
			8.35%	10.99%	0.83%	1.93%
	SA	0.24850	0.27240	**<u>0.27934</u>**	0.24881	0.25431
			9.62%	**<u>12.41%</u>**	0.13%	2.34%
United States	U	3.41673	4.11576	3.95774	4.03864	**<u>4.07116</u>**
			20.46%	15.83%	18.20%	**<u>19.15%</u>**
	SA	3.26239	4.16751	**<u>4.04854</u>**	4.01225	4.02966
			27.74%	**<u>24.10%</u>**	22.98%	23.52%
Uruguay	U	1.06546	0.94425	0.98781	**<u>1.02420</u>**	1.11378
			-11.38%	-7.29%	**<u>-3.87%</u>**	4.54%

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
Percentage difference values are in relation to the base model.
Bold and underscored values indicate alternative models that are a preferred model.

Comparing the out-of-sample performance of 2021 against 2020, Canada, Italy, Mexico, South Korea, the United States, and Uruguay see a model get more accurate. In particular, the countries that have their preferred model improve in 2021 versus 2020 are Canada, Italy, and the United States, the last being the only of the three with a preferred alternative model. Figure 5.1 plots the AIC of the preferred models and their respective RMSEs for 2020 and 2021 forecasts.

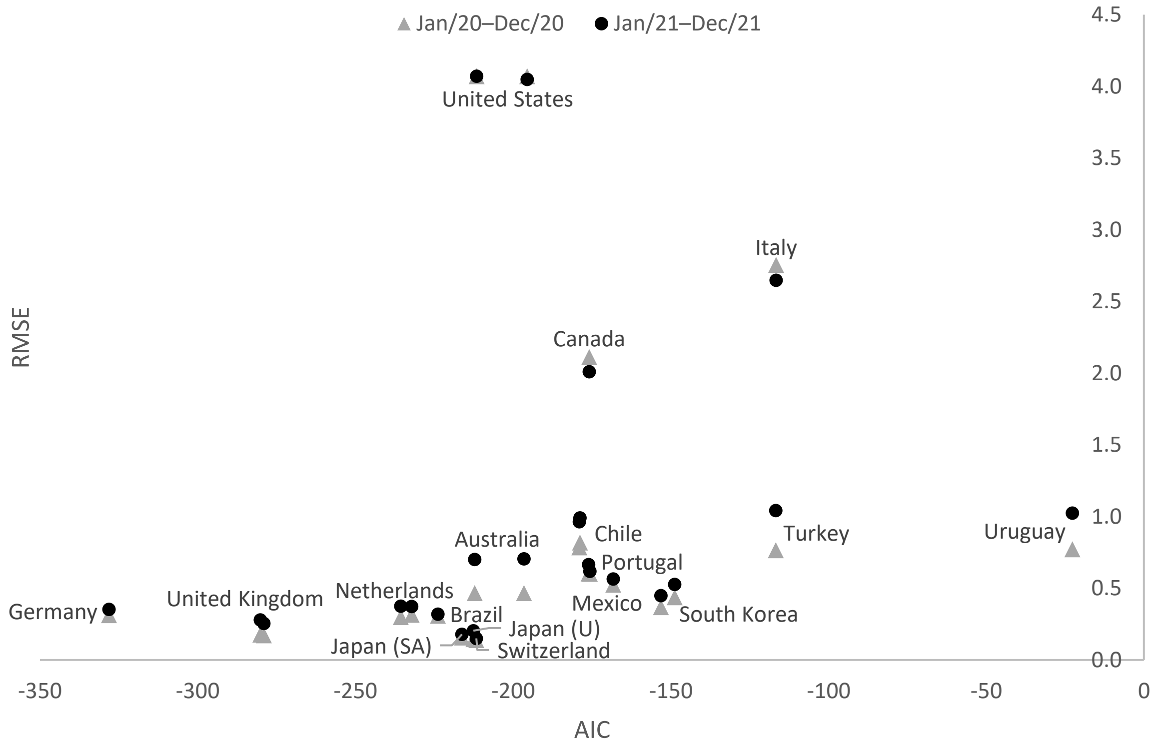

Figure 5.1 – AIC and RMSE values for 2020 and 2021 of the preferred models, Scenario A

With more than half of the countries not having search query data as part of their preferred model, it suggests that search query models do not impart much information that is not already contained in a simple autoregressive component. To make matters worse, only in two cases the preferred alternative models have better out-of-sample performance than the base model. Even though half the countries have at least one alternative model outperform the base in 2020 or 2021, search query data may not be a reliable solution to improve nowcasting accuracy. This can be further seen in the general decrease in forecasting power in 2021 versus 2020, which also puts into question the robustness of the data as predictor variables.

5.2. Scenario B: 2016–2020

Given that in this scenario the training window includes the spikes in unemployment often seen at the beginning of the Covid-19 pandemic, it is reasonable to expect a larger number of countries with alternative models as best performing than in Scenario A, as a purely autoregressive model is poor at forecasting outliers.

Indeed, eight out of 12 countries with unadjusted time series and seven out of 12 countries with seasonally adjusted time series had alternative models outperform the base model without GT data, for a total of 10 countries: Canada, Chile, Germany, Italy, Japan, the Netherlands, Portugal, the United Kingdom, the United States, and Uruguay. In those countries, models with AR(1) components and GT data, with or without lagged categories, had the lowest AIC values in 11 out of 15 cases, and lower AIC values than the base model in 14 out of 15 cases. Table 5.4 presents the AIC values for each model.

Table 5.4 – AIC values per model per variable per country, Scenario B

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	-167.68	-165.46	-164.57	-164.44	-165.00
	SA	-166.97	-165.06	-163.65	-163.72	-163.35
Brazil	U	-191.61	-149.59	-147.75	-189.93	-187.76
Canada	U	-22.41	-14.69	-78.14	-37.52	**-123.71**
Chile	U	-131.43	-102.15	-110.42	-127.57	**-145.26**
	SA	-134.47	-107.33	-115.62	-130.63	**-148.69**
Germany	U	-234.21	-229.96	-242.08	-230.36	**-246.62**
Italy	SA	-83.27	-111.38	-109.39	**-120.22**	-116.28
Japan	U	-215.15	-206.85	-206.62	**-216.42**	-215.44
	SA	-215.13	-213.60	-212.84	**-217.77**	-215.93
Mexico	SA	-138.48	-117.65	-115.64	-135.83	-134.35
Netherlands	U	-205.40	-204.18	-200.86	-202.72	-199.53
	SA	-205.23	**-206.45**	-203.08	-204.50	-201.14
Portugal	U	-133.68	-131.17	**-137.72**	-130.49	-136.65
	SA	-134.93	-132.39	**-138.23**	-131.34	-136.83
South Korea	U	-131.52	-118.98	-115.15	-128.56	-124.62
	SA	-140.35	-132.19	-128.60	-138.33	-134.33
Switzerland	SA	-209.13	-186.06	-182.86	-205.50	-203.30
Turkey	SA	-82.29	-72.42	-71.34	-80.56	-78.38
United Kingdom	U	-261.38	-257.39	-261.36	-257.65	**-261.90**
	SA	-261.25	-259.61	-263.16	-258.27	**-261.97**
United States	U	49.93	-25.16	-43.81	-30.45	**-65.32**
	SA	50.19	-24.25	-43.04	-29.06	**-64.06**
Uruguay	U	-15.45	-7.74	-4.88	**-19.09**	-15.25

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
Bold and underscored values indicate alternative models that are a preferred model.

Out-of-sample forecasts for 2021 are also better than in Scenario A. The base model has the lowest RMSE in only three countries: Brazil, Japan, and the Netherlands, the last being one of the countries with an alternative model surpassing the base in AIC. Of the other thirteen countries, only the United Kingdom has just one alternative model with lower error than the base; meanwhile, all alternative models outperform the base model for South Korea, the United States, and Uruguay. The best improvement is seen in the seasonally adjusted time series for the United States with a drop of 62,3% in RMSE compared to the base model when using GT data with lagged categories and an AR(1) component.

When looking only at the preferred models per variable per country, 10 out of 24 outperform the base models while 5 underperform against them, with the percentage difference in RMSE ranging from -62,3% to -4,0% and +1,5% to +29,3% respectively. Figure 5.2 plots the AIC of the preferred models and their respective RMSEs for 2020 and 2021 forecasts, and Table 5.5 presents all out-of-sample forecast results for 2021.

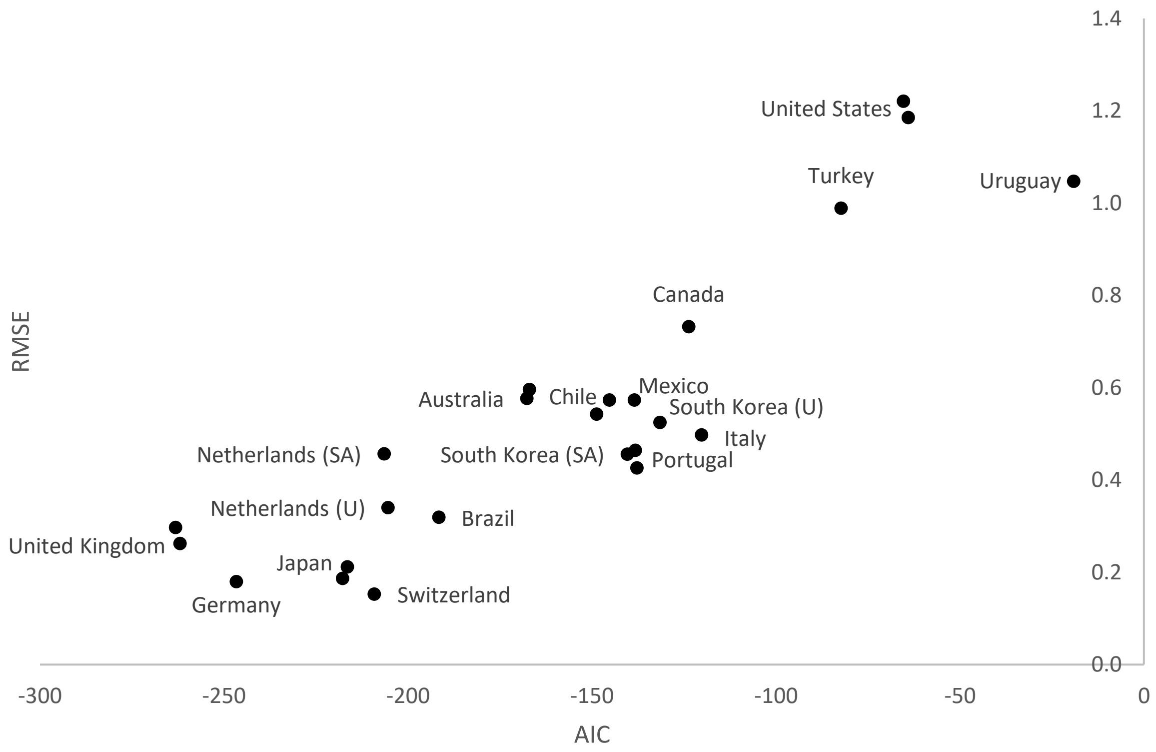

Figure 5.2 – AIC and RMSE values for 2021 of the preferred models, Scenario B

Table 5.5 – Out-of-sample forecast RMSE values and percentage differences for 2021, Scenario B

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	0.57642	0.60310	0.58848	0.56031	0.52376
			29.06%	25.93%	19.91%	12.09%
	SA	0.59580	0.61586	0.60052	0.58119	0.54731
			31.81%	28.53%	24.39%	17.14%
Brazil	U	0.31901	0.65181	0.64579	0.32765	0.32206
			111.13%	109.18%	6.13%	4.32%
Canada	U	1.53422	1.62754	0.99586	1.21164	**<u>0.73224</u>**
			-23.00%	-52.89%	-42.68%	**<u>-65.36%</u>**
Chile	U	0.64865	1.02168	1.04790	0.64939	**<u>0.57271</u>**

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
			24.50%	27.70%	-20.86%	**<u>-30.21%</u>**
	SA	0.62036	0.99149	1.01037	0.61401	**<u>0.54275</u>**
			25.99%	28.39%	-21.98%	**<u>-31.03%</u>**
Germany	U	0.29249	0.31276	0.23685	0.29117	**<u>0.17960</u>**
			0.16%	-24.14%	-6.75%	**<u>-42.48%</u>**
Italy	SA	0.87485	0.38721	0.42414	**<u>0.49770</u>**	0.49429
			-58.81%	-54.88%	**<u>-47.06%</u>**	-47.42%
Japan	U	0.20656	0.21932	0.24769	**<u>0.21183</u>**	0.23429
			52.91%	72.69%	**<u>47.69%</u>**	63.35%
	SA	0.18359	0.19177	0.21013	**<u>0.18630</u>**	0.20218
			19.38%	30.81%	**<u>15.98%</u>**	25.86%
Mexico	SA	0.57279	0.55165	0.57220	0.56889	0.58640
			5.21%	9.13%	8.49%	11.83%
Netherlands	U	0.33971	0.39808	0.36674	0.38872	0.35175
			26.56%	16.60%	23.59%	11.83%
	SA	0.35287	**<u>0.45624</u>**	0.43432	0.45518	0.43263
			<u>52.62%</u>	45.29%	52.27%	44.73%
Portugal	U	0.55916	0.55789	**<u>0.42579</u>**	0.53686	0.41498
			-2.01%	**<u>-25.21%</u>**	-5.71%	-27.11%
	SA	0.58790	0.58865	**<u>0.46397</u>**	0.57013	0.45418
			6.46%	**<u>-16.09%</u>**	3.11%	-17.86%
South Korea	U	0.52426	0.45848	0.45894	0.52203	0.52000
			31.87%	32.00%	50.15%	49.56%
	SA	0.45577	0.42055	0.42900	0.45457	0.45475
			45.78%	48.71%	57.57%	57.64%
Switzerland	SA	0.15228	0.23391	0.23125	0.14400	0.13403
			67.76%	65.84%	3.27%	-3.88%
Turkey	SA	0.98884	0.94397	0.92488	1.05171	1.01818
			23.07%	20.58%	37.11%	32.74%
United Kingdom	U	0.24321	0.28529	0.30061	0.24122	**<u>0.26242</u>**
			65.77%	74.67%	40.16%	**<u>52.48%</u>**
	SA	0.26007	0.28228	0.29711	0.25876	**<u>0.27594</u>**
			64.68%	73.33%	50.96%	**<u>60.98%</u>**
United States	U	3.17296	1.64167	1.39472	1.54176	**<u>1.22023</u>**
			-52.37%	-59.53%	-55.26%	**<u>-64.59%</u>**
	SA	3.14488	1.60175	1.37722	1.52074	**<u>1.18515</u>**
			-51.92%	-58.66%	-54.35%	**<u>-64.43%</u>**
Uruguay	U	1.09094	0.96543	0.94779	**<u>1.04696</u>**	1.06386
			29.91%	27.53%	**<u>40.88%</u>**	43.15%

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
Percentage difference values are in relation to the base model.
Bold and underscored values indicate alternative models that are a preferred model.

Shifting the training window as to include 2020 data has a very positive impact on the overall performance of alternative models, be it during training—from seven to 10 countries with preferred alternative models—or out-of-sample forecasting in 2021—from nine to 13 countries with alternative models outperforming the base.

A significant improvement is seen when focusing on the performance of preferred models in 2021: only one model has lower RMSE than the base in Scenario A, while in Scenario B that number jumps to ten. This improvement can also be seen in the percentage difference in RMSE as the lowest decrease in error in Scenario B (-4,0%) is already higher than the single decrease in Scenario A (-3,9%).

Furthermore, when comparing the out-of-sample forecast performance for 2021 between models from Scenario A and Scenario B, at least one alternative model from the latter has lower RMSE than its equivalent in the former in 19 out of 24 series; in 11 of those, all Scenario B alternative models outperform their Scenario A equivalents. The only countries with Scenario A models that beat their Scenario B equivalents are Japan, the United Kingdom, and Uruguay. The percentage difference in RMSE ranges from -85,85% to +12,05%. Table 5.6 presents the RMSE value difference and percentage difference between Scenario A and Scenario B equivalent models for 2021.

Table 5.6 – Out-of-sample forecast RMSE differences and percentage differences between Scenario B and Scenario A equivalent models for 2021.

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Australia	U	-0.12876	-0.00625	-0.07581	-0.15109	-0.24664
		-18.26%	-1.03%	-11.41%	-21.24%	-32.01%
	SA	-0.10481	-0.01157	-0.08520	-0.11540	-0.20445
		-14.96%	-1.84%	-12.43%	-16.57%	-27.20%
Brazil	U	-0.00159	-0.19407	-0.25448	-0.11332	-0.13476
		-0.50%	-22.94%	-28.27%	-25.70%	-29.50%
Canada	U	-0.47725	-0.16617	-0.77108	-0.81237	-1.30114
		-23.73%	-9.26%	-43.64%	-40.14%	-63.99%
Chile	U	-0.34365	0.04188	0.07084	-0.35226	-0.43177
		-34.63%	4.27%	7.25%	-35.17%	-42.98%
	SA	-0.34426	0.04745	0.07095	-0.36138	-0.43396
		-35.69%	5.03%	7.55%	-37.05%	-44.43%
Germany	U	-0.05987	-0.00009	-0.07436	-0.06292	-0.17203
		-16.99%	-0.03%	-23.89%	-17.77%	-48.92%
Italy	SA	-0.16652	-1.31439	-2.57398	-0.56746	-2.15496
		-15.99%	-77.24%	-85.85%	-53.27%	-81.34%
Japan	U	0.00844	0.00398	0.01779	0.00754	0.02031
		4.26%	1.85%	7.74%	3.69%	9.49%
	SA	0.01097	0.00505	0.01366	0.00748	0.01896
		6.35%	2.70%	6.95%	4.18%	10.35%
Mexico	SA	0.00702	0.00629	0.02726	0.00077	0.01406
		1.24%	1.15%	5.00%	0.13%	2.46%
Netherlands	U	-0.03412	0.03605	0.03944	-0.00146	-0.00878

Country	Variable	Model 1	Model 2	Model 3	Model 4	Model 5
		-9.13%	9.96%	12.05%	-0.37%	-2.43%
	SA	-0.02207	0.04418	0.03613	0.03233	0.01414
		-5.89%	10.72%	9.07%	7.65%	3.38%
Portugal	U	-0.01047	-0.06069	-0.21265	-0.07151	-0.20388
		-1.84%	-9.81%	-33.31%	-11.75%	-32.94%
	SA	-0.00747	-0.07767	-0.25374	-0.08711	-0.24422
		-1.25%	-11.66%	-35.35%	-13.25%	-34.97%
South Korea	U	0.00250	-0.01248	-0.01614	-0.00511	-0.01721
		0.48%	-2.65%	-3.40%	-0.97%	-3.20%
	SA	0.01027	0.00390	-0.00238	0.00593	0.00370
		2.30%	0.94%	-0.55%	1.32%	0.82%
Switzerland	SA	0.00157	0.00469	-0.03421	-0.04545	-0.10326
		1.04%	2.05%	-12.89%	-23.99%	-43.52%
Turkey	SA	-0.05424	0.04061	-0.2476	-0.00746	-0.26461
		-5.20%	4.50%	-21.12%	-0.70%	-20.63%
United Kingdom	U	-0.01214	0.00862	0.01718	-0.01625	0.00214
		-4.75%	3.12%	6.06%	-6.31%	0.82%
	SA	0.01157	0.00988	0.01778	0.00995	0.02163
		4.66%	3.63%	6.36%	4.00%	8.50%
United States	U	-0.24378	-2.47409	-2.56301	-2.49688	-2.85093
		-7.13%	-60.11%	-64.76%	-61.82%	-70.03%
	SA	-0.11751	-2.56577	-2.67132	-2.49151	-2.84451
		-3.60%	-61.57%	-65.98%	-62.10%	-70.59%
Uruguay	U	0.02548	0.02118	-0.04002	0.02276	-0.04992
		2.39%	2.24%	-4.05%	2.22%	-4.48%

U and SA indicate unadjusted series and seasonally adjusted series respectively.
Model numbering follows the numbering from the methodology section.
All values are in relation to the Scenario A equivalent model.

While these results are promising, they come at the expense of incorporating stress data into the training. It is still unclear what drives the improvements seen; just as they may come from the models being trained on the same stress scenario they are forecasting, it may also be the case that search query data is growing in forecasting power as more individuals search for and apply for jobs through the Internet.

6. CONCLUSION

It is impossible to outright dismiss the notion of ever adopting search query data into forecasting models, if not for the impressive results seen on occasion, then for the undisputed potential behind practical, high-frequency data that sources from millions of individuals and is publicly available at virtually no cost. However, it is still nebulous where to best put such data to use. Scenario A shows that the alternative models in almost half of the countries have less robustness to shock than an autoregressive component, so using this type of data as a detector to foresee a stress scenario seems ill-advised. Scenario B, on the other hand, showcases the potential of search queries to enhance forecasts, but having to train on a stress scenario to better forecast the very same stress scenario makes no assurances that the increase in forecast power holds up should there be a different stress scenario just over the horizon.

Overall, just as seen in the literature, there are situations in which search query data may improve unemployment rate one-month forecasts. At the same time, search query data do not provide an off-the-shelf solution. Australia, Switzerland, and the three not high-income economies included in the study—Brazil, Mexico, and Turkey—do not have a single alternative model with lower AIC than the base model in either scenario; Switzerland has no alternative model produce more accurate forecasts than the base model in Scenario A, while for Brazil the same happens in both scenarios.

There is a light at the end of the tunnel, however. As Internet access becomes more widespread, as more people change their job-seeking habits over to the Web, as search query categorization grows more accurate, it stands to reason that these types of data grow in forecasting power. Furthermore, the literature, including the study, focuses on using monthly query data to forecast monthly unemployment rate, even though the same online data is also available weekly or even daily, at the cost of time and ease of collection and aggregation. It may be the case that a mixed-frequency modelling approach such as seen in Ghysels et al. (2016) is better suited to this type of data. It remains to be seen if better model methodology or time will prove to be enough to realize the potential of this source of data in in-production models.

BIBLIOGRAPHICAL REFERENCES

Artola, C., Pinto, F., & de Pedraza García, P. (2015). Can internet searches forecast tourism inflows?. *International Journal of Manpower*.

Bantis, E., Clements, M. P., & Urquhart, A. (2021). Forecasting GDP Growth Rates Using Google Trends in the United States and Brazil. *Available at SSRN 3860194*.

Barreira, N., Godinho, P., & Melo, P. (2013). Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends. *NETNOMICS: Economic Research and Electronic Networking*, *14*(3), 129-165.

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

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google Trends in an emerging market. *Journal of Forecasting*, *32*(4), 289-298.

Choi, H., & Varian, H. (2009). Predicting initial claims for unemployment benefits. *Google Inc*, *1*, 1-5.

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

Damásio, B., & Mendonça, S. (2019). Modelling insurgent-incumbent dynamics: Vector autoregressions, multivariate Markov chains, and the nature of technological competition. *Applied Economics Letters*, *26*(10), 843-849.

Damásio, B., & Mendonça, S. (2022). Leader-follower dynamics in real historical time: a Markovian test of non-linear causality between sail and steam (co-) development. *Applied Economics*, 1-11.

Damásio, B., & Nicolau, J. (2014). Combining a regression model with a multivariate Markov chain in a forecasting problem. *Statistics & Probability Letters*, *90*, 108-113.

Damásio, B., & Nicolau, J. (2020). Time inhomogeneous multivariate Markov chains: detecting and testing multiple structural breaks occurring at unknown dates.

D'Amuri, F., & Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting, 33*(4), 801-816.

Elshendy, M., Colladon, A. F., Battistoni, E., & Gloor, P. A. (2018). Using four different online media sources to forecast the crude oil price. *Journal of Information Science, 44*(3), 408-421.

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment?. *Economic Modelling*, *30*, 117-125.

Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software*, *72*(1), 1-35.

Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing, 285*, 188-195.

Kandula, S., & Shaman, J. (2019). Reappraising the utility of Google flu trends. *PLoS computational biology*, *15*(8), e1007258.

Kristoufek, L., Moat, H. S., & Preis, T. (2016). Estimating suicide occurrence statistics using Google Trends. *EPJ Data Science, 5*, 1-12.

Kuzin, V., & Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting, 27*(2), 529-542.

Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics, 72*(4), 518-550.

Medeiros, M., & Pires, H. (2021). The Proper Use of Google Trends in Forecasting Models. *arXiv preprint arXiv:2104.03065*.

Mulero, R., & García-Hiernaux, A. (2021). Forecasting Spanish unemployment with Google Trends and dimension reduction techniques. *SERIEs*, 1-21.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change, 130*, 114-122.

Nagao, S., Takeda, F., & Tanaka, R. (2019). Nowcasting of the US unemployment rate using Google Trends. *Finance Research Letters, 30*, 103-109.

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of google trends in health care research: a systematic review. *PloS one*, *9*(10), e109583.

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013). Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS computational biology*, *9*(10), e1003256.

Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008). Using internet searches for influenza surveillance. *Clinical infectious diseases*, *47*(11), 1443-1448.

Salisu, A. A., Ogbonna, A. E., & Adewuyi, A. (2020). Google trends and the predictability of precious metals. *Resources Policy*, *65*, 101542.

Schaer, O., Kourentzes, N., & Fildes, R. (2019). Demand forecasting with user-generated online information. *International Journal of Forecasting, 35*(1), 197-212.

Silva, E. S., Hassani, H., Madsen, D. Ø., & Gee, L. (2019). Googling fashion: forecasting fashion consumer behaviour using google trends. *Social Sciences, 8*(4), 111.

Siliverstovs, B., & Wochner, D. S. (2018). Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. *Journal of Economic Behavior & Organization, 145*, 1-23.

Vicente, M. R., López-Menéndez, A. J., & Pérez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing?. *Technological Forecasting and Social Change, 92*, 132-139.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of forecasting*, *30*(6), 565-578.

Vosen, S., & Schmidt, T. (2012). A monthly consumption indicator for Germany based on Internet search query data. *Applied Economics Letters, 19*(7), 683-687.

Weng, B., Lu, L., Wang, X., Megahed, F. M., & Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications, 112*, 258-273.

Woo, J., & Owen, A. L. (2019). Forecasting private consumption with Google Trends data. *Journal of Forecasting, 38*(2), 81-91.

APPENDIX A: PAST AND CURRENT GOOGLE TRENDS SAMPLING

As all GT data is based on internal samples from the relevant queries rather than its entire population, it becomes pivotal to know when the samples change to expedite data collection and address the inherent sampling error. While the specifics are not publicly known, a rule of thumb is that all requests in a period of 24 hours from the same macrogeographical location—about continental scale—as based on the requesting IP address use the same internal sample. To gather data from different samples, then, it is necessary to either wait 24 hours or have access to IP addresses from different locations.

Past behavior was that data requests from the same sample produced the exact same time series, such that one easy way to automate data collection was to only accept a new observation if there were any different values in the entire series. On February 16th, 2022, however, this behavior was changed so that even successive data requests differ from the previous ones, albeit by only a few values. The first step of the data collection looks at interest over time on the Arts & Entertainment category in the United States to determine whether to carry on requesting the relevant data. Looking exclusively at the series pertaining to this first step, Figure A.1 plots the mean values from nine random series collected before February 16th, 2022, with the 10th and 90th percentiles shaded in gray, while Figure A.2 does so for the nine series collected past February 16th, 2022 before any behavior change was noticed.

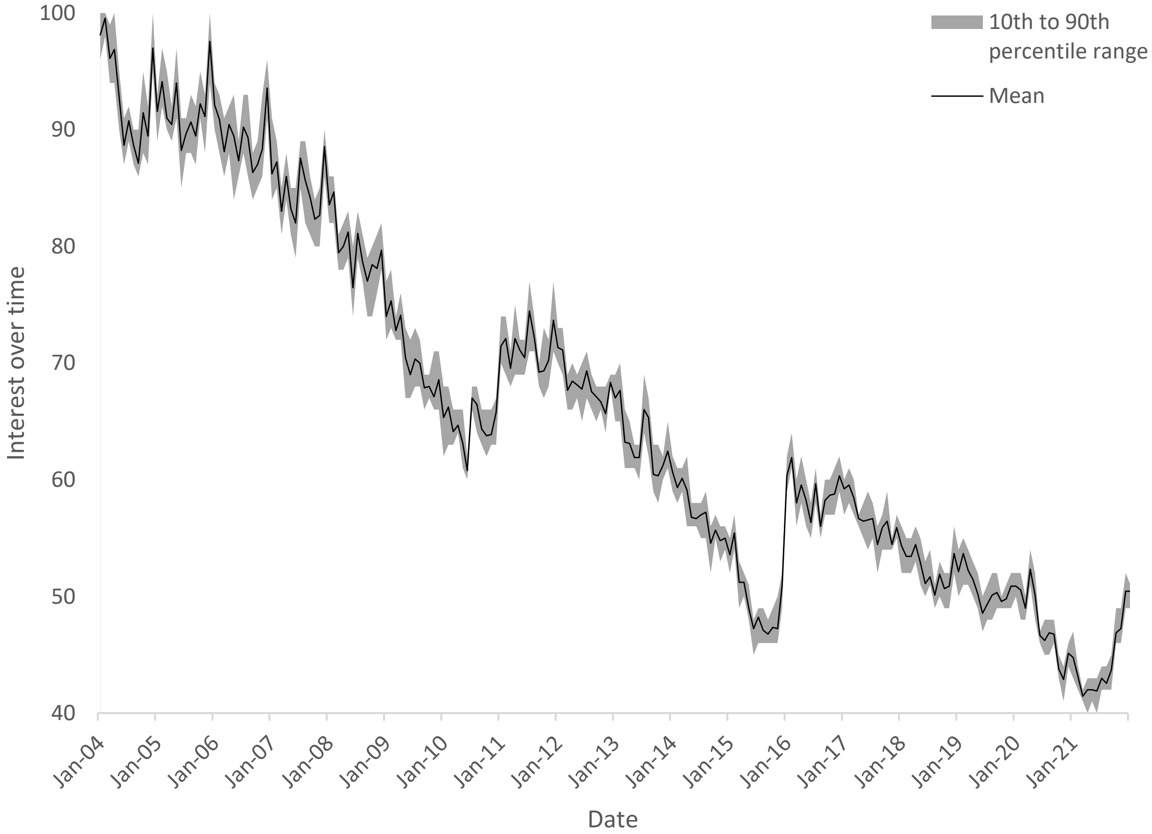

Figure A.3 – Mean, 10th and 90th percentile values for Arts & Entertainment in the United States under past behavior, n=9

Figure A.4 – Mean, 10th and 90th percentile values for Arts & Entertainment in the United States under current behavior before noticing any change, n=9

There is no month in which the 10th and 90th percentiles are equal for the nine random series collected before the change in behavior. In contrast, the number jumps to 162 (75%) when looking at the nine series naively collected past that date, before the change was noticed.

While the naïve approach of accepting any time series that is not exactly equal to any of the previously collected ones do not seem to work any longer, it is important to note that the successive series get more distinct from one another the longer the delay is between requests, approaching past behavior as the delay grows closer to 24 hours. Thus, one way to approach automated data collection with the current behavior is to set a threshold of minimum different values to accept a new observation.

For the study, such threshold is set to 108, half the number of months in the data request. After finishing data collection, the data was pruned down to observations with a minimum of 144 different values to another observation—two thirds of the number of months—resulting in a total count of 79 observations. Looking only at Arts & Entertainment queries in the United States, Figure A.3 plots the mean values from the selected 79 series, with the 10th and 90th percentiles shaded in gray.

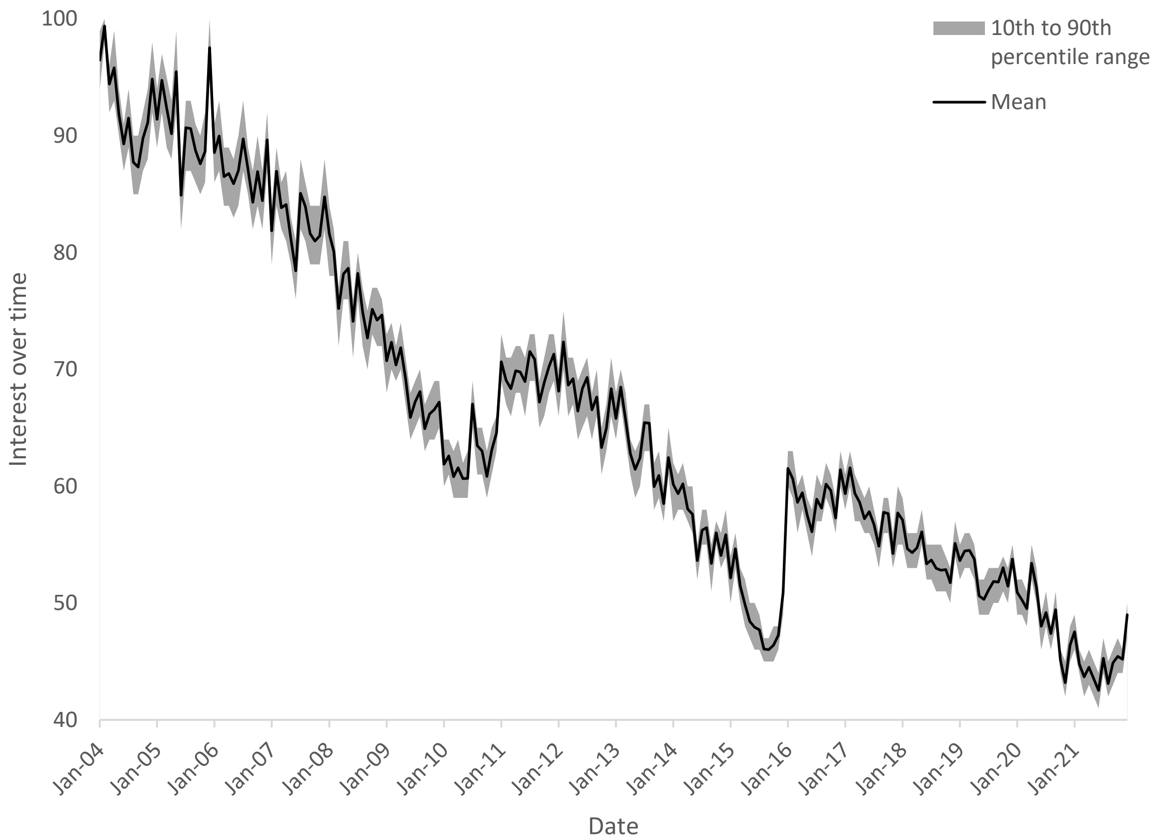

Figure A.5 – Mean, 10th and 90th percentile values for Arts & Entertainment in the United States under current behavior after pruning, n=79