

Search for a New Home: Refugee Stock and Google Search

International Migration Review

1-29

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01979183241275452

journals.sagepub.com/home/mrx



Ebru Sanliturk 

Max Planck Institute for Demographic Research, Rostock, Mecklenburg-Vorpommern, Germany

Francesco C. Billari

*Bocconi University, Milan, Lombardy, Italy
Dondena Research Center on Social Dynamics and Public Policy, Bocconi University, Milan, Lombardy, Italy*

Abstract

Following the assumption that trends of online queries may indicate intentions and help to predict human behavior, this study addresses the general issue of analyzing, nowcasting, and predicting migrant decisions through an analysis of Google search patterns in the case of Syrians in Turkey. Aiming to contribute to the literature on predicting migration patterns, we examine the relationship between Google search queries for province names in Turkey and the number of Syrians under temporary protection across provinces from January 2016 to December 2019 and demonstrate a positive and significant association. Then, we explore the predictive power of Google searches in predicting the stock of Syrians under temporary protection in Turkey across provinces. We exploit the alphabetical difference between Turkish and Arabic as the method of differentiation between host and migrant populations. Our findings indicate that Google searches can be good predictors for estimating refugee stocks, especially when traditional data are not available. They can also be helpful in forecasting the changing pattern of migrant stocks at frequent intervals, to which conventional socioeconomic indicators are less sensitive due to their less frequent reporting periods.

Corresponding Author:

Ebru Sanliturk, Max-Planck-Institute for Demographic Research, Rostock, Mecklenburg-Vorpommern, Germany.

Email: sanlituerk@demogr.mpg.de

Keywords

Syrian refugees, digital trace data, migrant stocks, forced displacement

Digital trace data gathered the attention of social scientists more than a decade ago as a new data source, and ever since, the analyses and interpretation of digital footprints acquired growing interest. Digital trace data are especially helpful when traditional data, such as official registries, census data, or national surveys, are lacking or data collection is burdensome and limited due to sensitivity and confidentiality issues. Data on forced displacement often suffer for the lack of official registries and limitations due to data protection. Digital trace data analysis thus contributes to forced displacement and refugee studies by presenting information on communities to which researchers have limited access and almost no real-time access.

Migration literature has shown that intentions to move, if and when they can be measured, help explain patterns of migration behavior (De Jong et al. 1985). Assuming that online search queries indicate intentions and help to predict human behavior, this study addresses the possibility of predicting the decisions to move in the context of forced displacement using online search data. We aim to shed light on this issue by exploring the potential of online searches to predict the moves of Syrians inside Turkey. We exploit the alphabetical difference between the local and migrant populations as a method of differentiation. We use this difference first to demonstrate the association between Google search data and the number (stock) of Syrians under temporary protection (hereafter SuTP) who reside in Turkish provinces. Then, we test the predictive power of online searches in estimating SuTP stocks across provinces in Turkey. By testing the potential of online searches as a predictor for refugee stocks across provinces and by exploiting the alphabetical differences between host and refugee populations as a method of differentiation, we contribute to the research on forced displacement as well as to the literature on the use of digital trace data in general and online search data in specific.

Background*Digital Trace Data*

The concepts of digital trace data and big data also referred to as online generated data, came to the attention of researchers in social sciences in the last two decades and paved the way for a growing literature in various fields. Being publicly available and easy to access, enabling real-time data collection, and providing opportunities to gain insight into hard-to-reach populations are the main advantages that digital trace data offer. At the same time, digital trace data also pose challenges, such as the problem of representativeness due to the lack of universal access to the internet and smartphones (Cesare et al. 2018). In response to the challenge of representativeness, the concepts of online populations and the digital divide emerged, addressed by

different methodologies using online data sources such as Twitter (Yildiz et al. 2017) and Facebook (Gil-Clavel and Zagheni 2019).

Migration literature also benefited from the increased use of digital trace data in research, as this new source of data offered new insights where traditional data and official records were not sufficient or timely. The geolocation information provided by most social media platforms is used as a proxy to examine international migration. Thus, social media platforms such as Facebook (Zagheni, Weber and Gummadi 2017), Twitter (Hawelka et al. 2014), and LinkedIn (State et al. 2014) became important data sources to monitor and interpret international migration flows. In the forced migration context, online trace data are shown as a useful tool to nowcast the displacement in Ukraine following the Russian invasion of Ukraine in 2022 (Leasure et al. 2023).

The main source of digital trace data in this study, online search trends and query popularity attracted the attention of researchers as a possible estimator to predict future tendencies and events (*forecasting*) and to interpret the present, defined as *now-casting*. Examples of such research first emerged in economics and epidemiology; to predict certain macroeconomic indicators (Ettredge, Gerdes and Karuga 2005; Askitas and Zimmermann 2009; Choi and Varian 2009; Guzman 2011), and infectious disease outbreaks (Brownstein, Freifeld and Madoff 2009; Ginsberg et al. 2009; Pelat et al. 2009) respectively. Furthermore, demography literature provided evidence for an association between internet search patterns and abortions (Reis and Brownstein 2010), fertility (Billari, D'Amuri and Marcucci 2016; Wilde, Chen and Lohmann 2020), and causes of mortality (McCarthy 2010; Song et al. 2014; Chang et al. 2015; Solano et al. 2016; Ricketts and Silva 2017).

Online search data has undoubtedly attracted the attention of migration studies. Two strands of research appear in migration studies that use online search behavior as a predictor. The first strand considers online search data as a proxy for migration and mobility trends, while the second focuses on public opinion on immigration and migrants (Mellon 2014; Lorenz et al. 2022) as well as among migrants (Bail, Merhout and Ding 2018; Chykina and Crabtree 2018). This study builds on and contributes to the first strand that uses online search data as a measure of intention to move to analyze and predict migration patterns.

Google Trends data are influenced by micro- and macro-level factors, as shown in the case of Puerto Rican migration to the US (Vicéns-Feliberty and Ricketts 2016). Migration-related online searches have a positive and significant association with international migration flows in the cases of Latin American migration to Spain (Wladyka 2017), migration to OECD countries (Böhme, Gröger and Stöhr 2020), migration to Switzerland from selected European countries (Wanner 2021), and Japanese migration to Europe (Leysen and Verhaeghe 2023). While these studies establish the predictive capacity of online search data in the migration context, Böhme, Gröger and Stöhr (2020) further provide evidence of their predictive performance in analyzing international migration flows by out-of-sample predictions and

showing that they outperform Gallup survey indicators for migration intentions. Building onto their findings, Google Trends data are proven to be useful predictors for short-term forecasting (now-casting) of Romanian migration to the UK (Avramescu and Wiśniowski 2021).

In the context of internal migration and mobility, online search data improve the strength of the statistical models to forecast US domestic migration (Lin, Cranshaw and Counts 2019) and interregional migration between Moscow and St. Petersburg (Fantazzini et al. 2021). Online search data are also used as a predictor to analyze how the COVID-19 pandemic has changed intentions to move and residential mobility preferences (Lei and Liu 2022). In the forced displacement context, Connor (2017) demonstrated the association between the online queries made in Arabic in Turkey for the word Greece and the asylum applications by Syrian refugees in Greece up until the EU-Turkey Deal in 2016. Yandex search data were also shown to present a timely tool for mobility as a proxy for the intention to move in the case of Russians leaving the country after the Russian invasion of Ukraine (Anastasiadou, Volgin and Leasure 2024).

This study builds on Connor's (2017) analysis of the case of Syrian refugees. From 2016 onwards, the movement of Syrian refugees was restricted to Turkey, which allows us to observe the movements of a large group of displaced people within boundaries. In contrast to Connor's (2017) strategy to exploit the language difference for country names in online inquiries, we exploit the difference in the alphabet, as the province names mostly remain the same over different languages. Exploiting the alphabetical difference in searching for province names, we analyze whether Google Trends can predict the stocks of refugees within a country.

Syrians under Temporary Protection in Turkey

The civil war in Syria, which broke out in 2011, escalated in the following years. Its effects extended well beyond the country's borders, especially in 2014–2015, as the refugee influx into the Mediterranean countries reached a peak. As a neighboring country, Turkey began receiving Syrian refugees with the start of the war, yet experienced the height of the crisis in 2014–2015. The European Union, through Greece and Italy, was also shaken by the Syrian refugee crisis. As the urgency of the situation became evident, the Turkish government became a part of the Syrian Regional Response Plan (SRRP) in cooperation with the UNCHR (Kirişçi and Ferris 2015) and signed the controversial readmission agreement with the EU (known as the EU-Turkey Deal) in 2015. The EU-Turkey Deal imposed Turkey to intercept refugees who wish to enter the EU borders through Turkey, to strengthen its capabilities to register, accommodate, and facilitate Syrian refugees, and to close the passageway to the EU to refugees through Turkey, which reflects the central policy of the EU (McMahon and Sigona 2021). The deal also conditioned the EU to accept one Syrian refugee to be resettled in the EU for each Syrian refugee being returned to

Turkey by the EU. The EU-Turkey Deal is important for this research as it marks the starting point of the period analyzed in this research, as the provincial-level data only became available after the conditions it placed on Turkey.

Understanding the settlement patterns of SuTP is important in the case of Turkey. Once the biometric registration of Syrian refugees after their arrival to the country is complete, they hold the status of temporary protection (hereafter TP), a status created to address the Syrian refugee crisis exclusively (Temporary Protection Regulation 2014). Less than 2% of the SuTP in Turkey live in camp environments, while the remaining majority live in urban areas. Although unregistered Syrian refugees certainly exist, the TP status enables access to free healthcare, education, and other forms of assistance through national and international initiatives. Access to public services is only possible in the province of registration; therefore, obtaining and maintaining the TP status is considerably beneficial. SuTP outside the camps can move to different provinces, provided they get a travel permit that can be issued for various reasons such as work, family visits, healthcare, and education by the Provincial Directorate of Migration Management (PDMM). After they move to the new province, they need to register in that province to continue to benefit from the rights and privileges of TP status. By the end of December 2019, when our empirical analysis ends, the total number of Syrian refugees under the temporary protection status was 3,576,370. Figures 6 and 7 in the Appendix show the change in the distribution of Syrian refugees across provinces by depicting the situation at the beginning of 2016 and the end of 2019, respectively, using Directorate General of Migration Management¹ (hereafter DGMM) data.

Data

This study explores the association between the official number of registered SuTP across provinces in Turkey and online search popularity for the province names in Turkey. The analysis period begins with the enactment of the EU-Turkey Deal in January 2016, when detailed regular statistics on Syrian refugees under temporary protection started to be published and ends on December 31, 2019. The two principal data sources, official records and Google Trends, are prepared for this time frame at the province level (81 provinces).

Official Data

The data source for the number of SuTP is the DGMM, the public body responsible for collecting and publishing information on international migration. In line

¹Original name of the institution is *İçişleri Bakanlığı Göç İdaresi Genel Müdürlüğü*

with the requirements of the EU-Turkey Deal, SuTP are registered at the province level, and the data have been publicly available since January 15, 2016.² While the DGMM publishes weekly data on the registered SuTP by province, the preceding data disappear on the public webpage with each update. We addressed this problem by using the Wayback Machine. The Wayback Machine captures and archives webpages through web crawls to save and retrieve the data lost for public access (Arora et al. 2016). Furthermore, we scanned the National Thesis Center³ database as the distribution of SuTP across provinces is occasionally provided as descriptive statistics in theses and dissertations (Arslan Erol 2018; Erkan 2018; Gök 2019; Önder 2019)⁴ on SuTP. We were able to build a data set of 122 weeks spanning from January 2016 to December 2019. Most of our data for the number of SuTP was obtained through our weekly downloads on the DGMM website and via Wayback Machine. For the data on province-level population and GDP per capita, we used the official end-of-year statistics reported by the Turkish Statistical Institute (hereafter TurkStat).

Google Trends Data

The online search popularity data source for the province names in Turkey is Google Trends. Google Trends provides a normalized score for the online search popularity of selected queries for the specific location and time frame rather than the sheer online search volume of the said queries. According to the guidelines provided by Google, the Google Trends data are computed by dividing each data point by the total searches in the specific location and time range to compare relative popularity, and the resulting numbers are then scaled on a range of 0 to 100 (Google 2022). The weekly Google Trends data correspond to the average daily search popularity in a week and is reported each Sunday.

Google Trends data come with certain limitations that need to be acknowledged. The online popularity score of a given query is not calculated based on the total amount of searches on Google but based on a sample with an unknown size, which the algorithm deems representative enough. Google Trends justifies this approach by pointing out that the actual volume of online

²Data on Syrians under temporary protection was available only at country level before that date.

³Full name; Council of Higher Education National Thesis Center (Yüksek Öğretim Kurumu Ulusal Tez Merkezi) accessible online via <https://tez.yok.gov.tr/UlusalTezMerkezi/giris.jsp>

⁴We included these descriptive statistics charts in our data sample only if they properly cited the same DGMM webpage source, included the table of statistics as published on the DGMM website unchanged and only if the week of the update is specified.

searches is over a few billion per day and calculations on the entire data for user-specified time periods would be too cumbersome. For each Google Trends query, the algorithm uses a different sample; therefore, reiterated queries are never based on the same sample. Google Trends output also does not allow the calculation of the number of total searches or any connection of online searches to internet users. In the migration context, this means that we cannot tell the number of people engaged in online information seeking, whether the same people do searches for multiple destinations or who turn the intention to move into action and who do not.

Migration literature using online search data has so far been more focused on labor migration from the perspective of both internal (Lin, Cranshaw and Counts 2019) and international migration (Wladyka 2017; Böhme, Gröger and Stöhr 2020; Avramescu and Wiśniowski 2021; Wanner 2021; Leysen and Verhaeghe 2023). These studies include an elaborate set of keywords to capture the online information search of potential migrants. The possible migration destination is taken as part of the online search query. Using a large set of keywords together also allows us to escape from the noise in the data or lack of data due to the Google Trends algorithm that deems the popularity of searches negligible for reporting.

In the context of forced displacement, however, we have less information about the keywords/phrases used by refugees and no chance of conducting a survey to determine them, especially while they are on the move. Therefore, we use only the province names in Turkey as the online search query, similar to the approach used by (Connor 2017). We exploit the alphabetical difference to distinguish the host and refugee populations. Province names rarely change in different languages, and if it does, it applies only to the large and well-known provinces. This approach allows us to observe the interest in the potential destination province and keeps the queries comparable between the hosts and refugees, as their information needs are expected to be different. The lack of Google Trends data, due to low interest and popularity, is less problematic in this approach. We consider too low interest as zero or having too many zero values in a time series not as a lack of capturing online search queries but rather as a lack of interest in the relevant place.

In our study, the specified query index is the names of provinces in Turkey in Arabic and Turkish letters.⁵ For the spelling of Turkish province names in Arabic, we used the version in the Wikipedia pages of each province in Arabic. We have also taken into account alternative spellings or misspellings for province names in Arabic letters and considered them together with the original/most common

⁵Turkish letters refer to the Latin alphabet with the special characters in the Turkish language (ç, ğ, ı, ö, ş, ü)

spelling.^{6,7} The keywords used for the Google Trends inquiry are reported in Table 2 in the Appendix.

The advantage of using alternative spellings, especially in our study where we do not have phrases for information gathering but the name of the potential destination (province) in Turkey, is to capture the popularity of the word better and with less noise in the data. We demonstrate this with the example of the province of Adana and using Google Trends' comparison feature. Google Trends allows up to five keywords to be compared in terms of popularity, and the comparison feature normalizes the results not only by the given time and location but also among the selected keywords. Taken separately, each query shows only the normalized score for that query for the given time and location. We visualize the alternative spelling approach in two steps. First, we look at the popularity of different spellings for Adana in Arabic letters as well as all spellings together as separate queries (Figure 8 in the Appendix). We then look at the four different spellings for Adana in Arabic letters and all spellings together comparatively (Figure 9 in the Appendix). The comparison, in which the results are normalized relative to each other, shows that using all possible spellings captures the popularity better than even the closest-performing single query (Figure 1).

The period of our empirical analysis is from January 1, 2016, to December 31, 2019. We set the time frame for the Google Trends data to begin on November 15, 2015, and end on December 31, 2019. The data for 2015 is used as the lagged search popularity and was not matched with the registry data of SuTP. For controls, we used Google Trends data for the same time period and query index in Syria, both in Arabic and Turkish letters, as well as the international spelling of province names without special characters⁸ in both Turkey and Syria. To test the potential bias of Google Trends sampling, we downloaded the data more than twice to see the difference between different samples but did not observe any significant change.

Data Preparation

Using the sources above, we created a unique panel data set covering four years, which includes 216 weeks of Google search data in six categories (Arabic letters in Turkey, Turkish letters in Turkey, English letters in Turkey, Arabic letters in

⁶Google Trends allows for different spellings to be considered together. If potential spellings are written as one query separated by + (OR) sign, Google Trends reports the combined results that include any of these spellings. e.g. Spelling1 + Spelling2

⁷To identify alternative spellings, we also used the Google Correlate. An online tool by Google that used to show the related online search queries for each given keyword. Google Correlate was shut down by Google at the end of 2019.

⁸International spelling refers to using English language characters instead of special characters (c, g, i, o, s, u respectively).

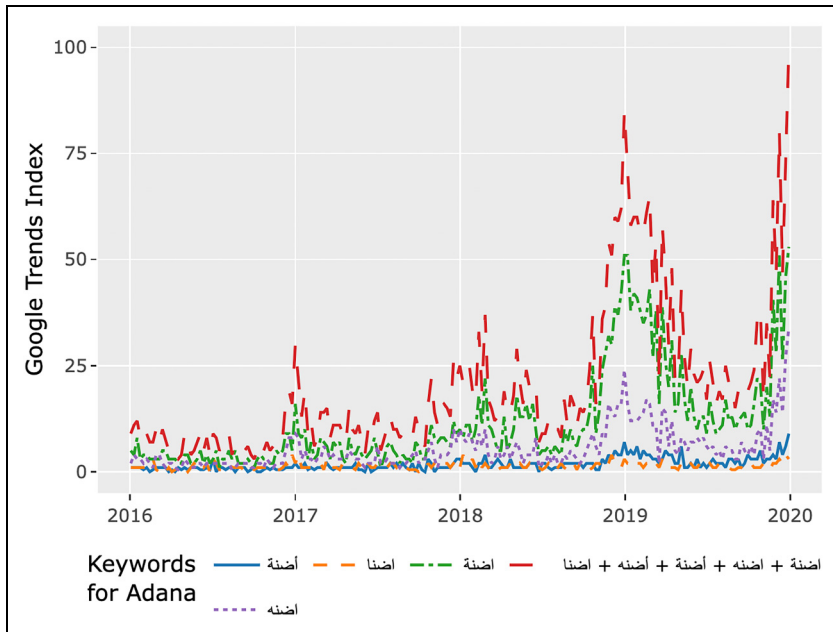


Figure 1. Comparison of alternative spellings for the province of Adana.

Note: If potential spellings are written as one query separated by + (OR) sign, Google Trends reports the combined results that include any of these spellings. The comparison tool of Google Trends reports the Google Trends Index for queries relative (normalized) to each other. Figure 1 shows the comparison of the Google Trends Index measuring the search popularity for four different spellings of the province Adana and the four different spellings combined as one parameter. The four different spellings are shown in blue and solid, yellow and short-dashed, green and long dash-dot, and purple and dotted lines. Combining all spellings of Adana in the query (shown in red and long-dashed line) provides the most extensive results relative to the single queries.

Syria, Turkish letters in Syria and English letters in Syria) and 122 updates for the official number of SuTP in 81 provinces. On this data set, we first calculated the main explanatory variable as the ratio of search popularity score for province names in Arabic letters over Turkish letters in Turkey ($search\ popularity\ in\ Arabic\ letters_{i,t} / search\ popularity\ in\ Turkish\ letters_{i,t}$). The variable is referred as the search popularity ratio, hereafter SPR. Introducing a search popularity ratio as the predictor is a novel approach of this paper to better understand the divergences in the online search behavior of the refugee population with respect to the local population. The reason for using SPR instead of separate search popularity scores for Arabic and Turkish spelling is thus to absorb the effects of potential local or seasonal popularity of provinces. If a province in Turkey becomes popular for any extraordinary events or incidents, the search popularity for the name of the province is expected to rise among both SuTP and Turkish citizens.

In the data preparation step, we calculated the time-lagged SPR variables and matched the official statistics for the number of SuTP across provinces with the Google Trends data (SPR). The official statistics of SuTP are published on weekdays, while Google Trends reports weekly data for each Sunday; therefore, the two data sources can be matched by week and not by day. As the SPR variable refers to the searches made in the same week of the update in the official registries for SuTP, this variable accounts for the nowcasting predictor in our model. As forecasting predictors, we calculated the six lagged variables with one-week intervals and for both the SPR variable and the search popularity for controls over 216 weeks. Once all the SPR and search popularity variables were set, we matched the dates of 122 updates of SuTP registry data with the dates of SPR and search popularity control variables by week. We then dropped the weeks that we could not match with the data on SuTP, and this way, we ensured that time lags were not lost or mismatched due to the shrinking sample size.

Our assumption in using Google Trends data to predict SuTP stocks is that they are an indicator of interest, and this interest might signal a move to the relevant province. However, Google Trends does not provide us with an absolute value but an index of popularity, and low interest, as interpreted by Google, is shown as zero. While this is challenging for doing research with online search data, we interpret that the lack of interest could also be meaningful for our predictions. We create three dummy variables that take the value one if the index value provided by Google Trends is 0. As the interest in Turkish provinces in Turkey using Turkish letters is never zero, we have three dummy variables for the absence of online search popularity; the first is for Turkish spelling with the location Syria, the second is for Arabic spelling with the location Syria and the last is for Arabic spelling with location Turkey.

Methods

The main aim of this study is to assess the association of online search popularity with the SuTP stock across provinces in Turkey and explore its predictive power on the latter. In this case study, we seek to understand how Google search data can be used to predict SuTP stocks. In contrast to the previous literature on this topic, we focus on refugee stocks instead of refugee flows. The reason for this choice relies on the characteristics of the selected case and the data availability. The SuTP registry data only gives the aggregate number of SuTP for each province and does not distinguish between new arrivals and relocations. As the relocations are not distinguished, the registry data also does not provide any information on the previous location of residence. The province-level SuTP registry data started being published after the EU-Turkey Deal and marked the beginning of our research period as 2016. At the beginning of 2016, Turkey was already hosting 2,540,784⁹ SuTP.

⁹As of 15.01.2016

Therefore, we must assume that the changes in the SuTP stock of provinces during our research period may be due to both new arrivals and relocations. In addition to the limits set by the registry data, the Google Trends data also favors an analysis of stocks rather than flows in this context. We set the location parameter for the Google Trends data as Turkey (country-level). While province-level Google Trends data would have provided us more insight into our research, they are more noisy and prone to an inflation of zero values, if available at all. It is, therefore, currently not possible to conduct an analysis of online interest in provinces at the province level or make assumptions on flows of SuTP between provinces. We need to assume that online searches that include the province name may be done while residing in the same province, making the Google Trends data we use in our research more suitable for analyzing stocks.

We define two hypotheses in our study. The first hypothesis (*H1*) suggests that online search popularity is positively associated with the number of SuTP in a given province and is a useful predictor for SuTP stocks. The second hypothesis (*H2*) proposes that the timing of online searches plays a role in predicting the association between online search behavior and the observed SuTP stock. To test these hypotheses, we use our unique panel data set and a fixed effects model, considering the advantages of fixed effects model to absorb the effect of features that vary at province-level and time-level and allow a better interpretation of the explanatory variables used in this analysis. Thus, the baseline model is as follows;

$$\ln(\text{Number of SuTP})_{i[tj]} = \beta_1 \frac{\text{search popularity in Arabic letters}_{i[tj]}}{\text{search popularity in Turkish letters}_{i[tj]}} + \alpha_i + \epsilon_{i[tj]} \quad (1)$$

In the specification above, the dependent variable is the natural logarithm of the number of SuTP in province i , at time tj , and the explanatory variable is the search popularity ratio (SPR) at time tj . We denote the time of the variable as tj , where j indicates the week of Google Trends data and t is the week in the registry data of SuTP. Thus, the time denotation of tj refers to any week for which Google search popularity and registry data of SuTP were matched (122 in total). The province-level fixed effects are denoted by α_i . The search query ratio refers to Google search popularity in Arabic letters for province i , at time t , over Google search popularity in Turkish letters for province i , at time t , in Turkey.

Next, to observe the role of the difference between the time of the Google search and the time SuTP stock appears on official data, the time-lags, we expand our model and add the lagged variables. While tj refers to the week for which the number of SuTP is matched with the same week of Google search query data, we denote the lagged variables as $tj-1$, $tj-2$, ... $tj-6$. Each lag refers to a one-week interval, and we introduce the lagged variables to the model, from one-week up to six-week lags in our data. Following the addition of lagged variables, the model becomes as

follows;

$$\begin{aligned} \ln(\text{Number of SuTP})_{i[tj]} = & \beta_1 \frac{\text{search popularity in Arabic letters}_{i[tj]}}{\text{search popularity in Turkish letters}_{i[tj]}} \\ & + \beta_2 \frac{\text{search popularity in Arabic letters}_{i[tj-1]}}{\text{search popularity in Turkish letters}_{i[tj-1]}} \\ & + \beta_3 \frac{\text{search popularity in Arabic letters}_{i[tj-2]}}{\text{search popularity in Turkish letters}_{i[tj-2]}} \\ & + \dots + \beta_7 \frac{\text{search popularity in Arabic letters}_{i[tj-6]}}{\text{search popularity in Turkish letters}_{i[tj-6]}} + \alpha_i + \epsilon_{i[tj]} \end{aligned} \quad (2)$$

Using online generated data as an explanatory variable raises questions about selection bias as the sample population is restricted to the portion of the population with internet access and high internet usage. Zagheni and Weber (2015) propose correcting the estimates by internet penetration rate to address this selection bias issue. This solution is problematic in the specific case of this study as the household and individual internet penetration rate in Turkey continued to increase during the time period of analysis. However, bearing in mind that these rates may not represent the SuTP in Turkey, we refrain from adding the internet penetration rate to the model to avoid introducing further bias.

We provide a closer look at internet penetration rates for the case study to justify the merits of studying online information-seeking patterns. During the period of study in Turkey, the share of households with internet access increased from 76.3% in 2016 to 88.3% in 2019, and the share of individual internet usage increased from 61.2% in 2016 to 75.3% in 2019 (TurkStat 2023). The region of residence, gender, and age contribute to the digital divide. Istanbul, with the highest population and SuTP population, also has the highest shares of households with internet access (89.1% to 95.6%) and individual internet usage (71.8% to 86.6%) throughout the period 2016–2019.¹⁰ On average, younger people and males have higher internet usage (TurkStat 2023).

Understanding the internet and smartphone usage of SuTP in Turkey is more difficult as there are no official statistics. Syrian refugees use the internet both during the journey to the destination and at the destination country. Smartphone and internet usage may make refugees vulnerable to risks from smugglers or states by rendering them easily trackable. However, evidence shows that Syrian refugees consider that the benefits of careful smartphone usage during the journey outweigh the risks

¹⁰Shares of households with internet access and individual internet usage are reported at the NUTS 1 (regional) level. Istanbul is the only province that is, at the same time, a NUTS 1 region.

(Gillespie, Osseiran and Cheesman 2018). In the case of SuTP in Turkey, one sees that smartphone ownership and internet usage are widespread. SuTP can acquire a phone number in Turkey using their IDs. While this may be an issue undocumented migrants want to avoid, we expect it to be negligible for the SuTP in this study, as the temporary protection grants them legal status.

Our knowledge of the smartphone ownership and internet usage of SuTP in Turkey comes from the surveys in the literature. Survey data shows that 33% of SuTP in Turkey never used the Internet in Syria, but only 10% are non-users in Turkey, and 75% are active Internet users¹¹ (Merisalo and Jauhiainen 2020). Smartphone ownership among SuTP in Turkey is estimated to be between 78% and 84% (Sunata 2017; Merisalo and Jauhiainen 2020). There is also evidence of a digital divide. SuTP residing in Western provinces and metropolitan areas generally have higher smartphone and internet usage (Narli 2018; Sunata 2017), and SuTP in camps, where freedom of movement is restricted, is specifically lower (Smets 2018). SuTP in camps are immobile but contribute to the SuTP stocks in the respective provinces. While we note that between 2016 and 2019, approximately 90% of the SuTP lived in urban areas, we acknowledge that the digital divide among the SuTP in Turkey may create selection bias.

Results

Statistical Analysis

In our analysis (Table 1), the dependent variable is the natural logarithm of the number of SuTP in each province. In the first column, the explanatory variable is the Arabic/Turkish SPR for searches made on the same week as the date of the SuTP registry (denoted as t_j) along with province-level fixed effects. In column 2, one-week, one-month, and six-week lags for the Arabic/Turkish SPR in Turkey are added to the equation. In the third column, we show the results for the Model 2 above. In column 4, controls are included in the regression equation: province-level population and dummy variables for the overall absence of online search popularity for province i at time t . We use the model in column 4 in our prediction exercises below (4.2) as the *paper model*. The last column shows the results for a more traditional approach in the literature to estimate migrant stocks, using province-level population, GDP,¹² and fixed-effects. We use this model as the *comparison model* in prediction exercises (4.2).

¹¹The study defines “active internet use” as using the internet every day or many times a week (Merisalo and Jauhiainen 2020).

¹²Official statistics on province-level population and GDP are published annually, for the previous year. Therefore, the last available end-of-year statistics are used in predictions for each year.

Table 1. Summary of Results.

	(1)	(2)	(3)	(4)	(5)
Arabic/Turkish SPR (t)	0.0928*** (0.0236)	0.0673*** (0.0151)	0.0563*** (0.0120)	-0.00133 (0.0198)	
Arabic/Turkish SPR (t-1)		0.0607*** (0.0164)	0.0459*** (0.0129)	0.0156* (0.00917)	
Arabic/Turkish SPR (t-2)			0.0530*** (0.0118)	0.0183* (0.0100)	
Arabic/Turkish SPR (t-3)			0.0348*** (0.0112)	0.00719 (0.00910)	
Arabic/Turkish SPR (t-4)		0.0658*** (0.0137)	0.0482*** (0.0103)	0.0185*** (0.00838)	
Arabic/Turkish SPR (t-5)			0.0540*** (0.0105)	0.0184*** (0.00858)	
Arabic/Turkish SPR (t-6)		0.0749*** (0.0166)	0.0590*** (0.0141)	0.0213* (0.0113)	
No SP in Turkish Letters (SYR)				-0.0568*** (0.0187)	
No SP in Arabic Letters (SYR)				-0.00463 (0.0140)	
No SP in Arabic Letters (TUR)				-0.0380 (0.0348)	
ln(Population)				6.190*** (1.193)	1.187 (0.971)
ln(GDP)					0.953*** (0.162)
Constant	8.168***	8.108***	8.080***	-73.79***	-23.27***

(continued)

Table 1. (continued)

	(1)	(2)	(3)	(4)	(5)
Observations	9,882 (0.00813)	9,882 (0.0196)	9,882 (0.0234)	9,882 (15.80)	9,882 (11.57)
Number of Provinces	81	81	81	81	81
R2 (within)	0.0143	0.0408	0.0529	0.212	0.322
R2 (overall)	0.0881	0.182	0.212	0.564	0.538

Note: Table 1 shows the results of the fixed-effects panel regression and demonstrates the association between the online search popularity of province names in Turkey and SuTP stock (measured in log form) across provinces. The robust standard errors are shown in parentheses. Significance levels are denoted as follows: *** $p < .01$, ** $p < .05$, * $p < .1$.

The analysis shows no support for a correlation between simultaneous online SPR and SuTP registry once the control variables are included. However, the analysis supports our *H1* and *H2* and confirms that there is a positive and significant association between the number of SuTP in a province and the online SPR for the name of the same province, as well as the timing of the online SP matters. The results suggest that an increase by one in the one-week prior SPR is correlated with a 1.6% increase in SuTP. The association becomes stronger when the time gap increases, as one-month and six-week time lags are associated with a 1.9% and 2.1% increase in the number of SuTP, respectively. However, an increase of about 2% in the number of SuTP in a given province in correlation with an increase by one in SPR should be considered carefully. As shown in Table 3 in the Appendix, the overall range of the SPR variable is 0 to 7.08, while the overall mean is 0.35. An increase in SPR by one is an unlikely event, and it is more likely that the association between the SPR and the number of SuTP in a province occurs in much smaller amounts. Controls for the absence of Google searches (No SP) are only significant for the absence of SP in Turkish letters in Syria. The absence of Google searches in Turkish letters in Syria is correlated with a 5.7% decrease in the number of SuTP. The result is intuitive as online searches for the provinces of another country using a foreign alphabet and special characters would be a clear sign of interest.

We implemented a series of robustness checks for our results. First, we replaced the SPR variable with SP in Arabic letters and SP in Turkish letters as two separate variables to test our approach of using the search popularity ratio as the explanatory variable and to disentangle their relationship with the dependent variable of the number of SuTP. Similar to the results of our initial models with controls, the simultaneous online SP variable was neither for the SP in Turkish letters nor for the SP in Arabic letters significant. The correlation between SP in Arabic letters and the number of SuTP is observed to be positive and statistically significant for a time lag of one, two, and four weeks.

Next, we replicated the model using the same queries and time period for the online SP variable, changing the online search location from Turkey to Syria. As the online search popularity for Turkish province names in Syria includes considerable zero values, we used the SP in Arabic letters and SP in Turkish letters separately instead of calculating SPR variables. The correlation between the online searches in Syria for Turkish province names in Turkish letters is positive and significant with the SuTP stock of these provinces for all time-lags (Table 4 in the Appendix). This result corroborates the strong negative correlation between the dummy variable for the lack of online searches in Turkish letters in Syria¹³ and SuTP stock in our main model. The correlation between the online searches in Syria for Turkish province names in Arabic letters is negatively correlated with the SuTP stock, significant only for a time lag of one week and six weeks. As the online search location is Syria,

¹³No SP in Turkish Letters (SYR)

we should bear in mind that the online search may be conducted not by the Syrian citizens who are about to move to the said province in Turkey but by the relatives and acquaintances of SuTP in Turkey.

Last, we replaced the SP in Turkish letters with SP in English letters, i.e., we removed the Turkish special characters from the online search query and repeated the analysis. The results for both Turkey and Syria, as the location of online searches, are observed to be quite similar to our initial results. Considering that Google Trends data are produced on a different and representative sample of all searches at every request, controlling for no Turkish characters and obtaining similar results also serves as a control for the same data from a different sample. Using English characters, however, created a difference for the analysis in Syria, and the association with the number of SuTP became stronger.

The Predictive Power of the Google Trends Model

After demonstrating the association between the online search popularity and the stocks of SuTP across Turkish provinces in the statistical analysis above, we proceed with two prediction exercises to test the accuracy of our model and online search indicators. In the first exercise, we seek to understand whether and how including online search indicators improve the prediction of SuTP stocks. We first look into how the time lags of the online searches impact the predictions, which addresses our H2. We follow by comparing our models' predictive performance with the predictive performance of a standard model without any online search indicators. In the second exercise, we compare the predictive performance of models with and without the online search indicators for random and non-random samples, where the non-random sample predicts the SuTP stocks for 2019, the final year in our dataset.

Exercise 1: k-Fold Cross-Validation. The k-fold cross-validation method enables the assessment of the predictive performance of a model by dividing the dataset into k random parts, training the model on the $k-1/k$ part of the dataset, and testing it on the $1/k$ part of the dataset. This process is repeated k times, and the root mean square error (RMSE) and R^2 values are reported for each repetition. We apply 5-fold cross-validation, in which the dataset is divided into five parts, and cross-validation is repeated five times for each model (80% / 20%). A comparison of RMSE values allows us to draw a comparative picture of the predictive performance of time lags for online search indicators and different models using/not using online search indicators.

Addressing the H2, we first assess the predictive power of different time lags in online searches with a cross-validation exercise. In Figure 2, the boxplot on the left-most side shows the dispersion of the RMSE from 5-fold cross-validation for a model that predicts the province-level SuTP stocks based on population, GDP, and time-invariant province fixed effects. Each subsequent boxplot shows the dispersion of

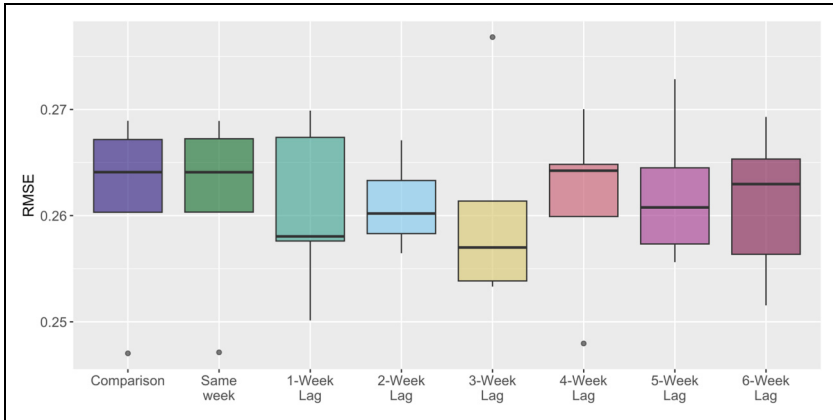


Figure 2. Predictive power of the time of online searches — RMSE comparison. *Note:* The figure demonstrates the RMSE values for different models from 5-fold cross-validation. The comparison model predicts the SuTP stock across provinces based on population, GDP, and province fixed effects. The following models add an online search indicator to the comparison model at different times to observe the change in the predictive power measured by RMSE. Each box represents the interquartile range of the RMSE values for the respective model, while the bold, black lines show the median value. The small black dots that may appear for some boxplots indicate the outliers.

the RMSE from 5-fold cross-validation for a model that adds an online search indicator to the comparison model, starting with the SPR of the same week up to the SPR from six weeks before. We observe that, on average, the inclusion of online searches from one to three weeks prior reduces the RMSE and thus improves the predictive power of the model, while the online searches at other times do not change the predictive power.

Next, we apply 5-fold cross-validation to four different statistical models to observe whether and how including online search (Google Trends-GT) indicators change the predictive power. The first model, named the *comparison model*, uses population and GDP as explanatory variables, as well as province-level fixed effects, and represents a more standard approach in the literature to predict migration. The *GT-only model* includes only the online search indicators (SPR and SP dummy variables) and province-level fixed effects, representing an approach for the most data-scarce situations. The *paper model* refers to the model shown in Table 1, column 4, which includes population and province-level fixed effects, and the online search indicators. The fourth model, shown as *GT+comparison* in Figure 3, is a merged model and uses all online search indicators, GDP, and population as explanatory variables, together with province-level fixed effects.

As Figure 3 demonstrates, the model that includes all variables, population, and GDP together with the GT indicators, on average, performs the best at predicting

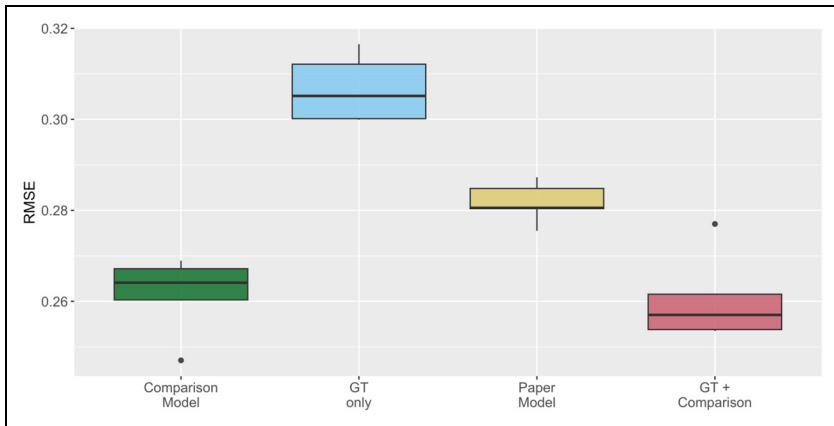


Figure 3. Predictive power of statistical models — RMSE comparison.

Note: The figure demonstrates the RMSE values for different models from 5-fold cross-validation to compare the predictive power of different statistical models. Each box represents the interquartile range of the RMSE values for the respective model, while the bold, black lines show the median value. The small black dots that may appear for some boxplots indicate the outliers.

the SuTP stock across provinces. However, the model without GT indicators predicts the SuTP stock better than the *paper* and *GT-only* models. The RMSE value of the *GT-only* model, despite being the highest comparatively, still poses an opportunity for the most data-scarce cases.

Exercise 2: Predictions on Held-out Data. In the second prediction exercise, we divide our dataset into two parts, train the models on the larger (train) portion of the dataset and predict the values in the smaller (test) portion of the dataset. We randomly determine the held-out data and then hold out 2019 (non-random data) from the dataset and repeat the prediction process. For the first round of predictions, the randomly selected held-out data consists of 2965 observations, referring to 30% of the dataset. To predict all observations in 2019, we hold out 30.3% of the dataset (2997 observations).

Figure 4 depicts a comparison of the RMSE values of different models (Table 5 in the Appendix) and shows their predictive performance on random and non-random data. The RMSE values resulting from the prediction of randomly selected observations show better predictive performance for the comparison model and the model that adds GT indicators to the comparison model. These results are in parallel with the results of the cross-validation exercise above (section “Exercise 1: k-Fold Cross-Validation”).

Prediction of SuTP stocks across provinces in the year 2019 (non-random sample), however, shows a contrasting picture. We observe the lowest RMSE

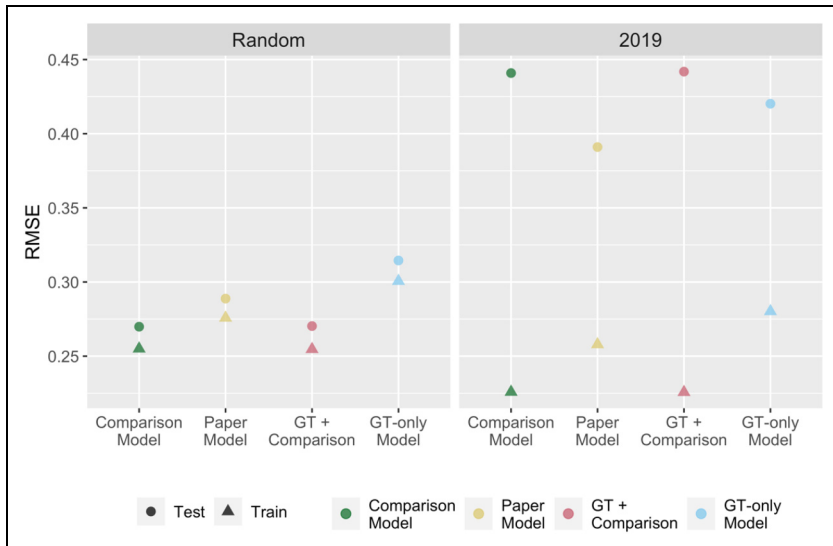


Figure 4. RMSE comparison of predictions on held-out data.

Note: The figure demonstrates the RMSE values to compare the predictive performance of different models on random and non-random (year 2019) test data. Circles show the RMSE for the test data, and the triangles show the RMSE for the train data. Smaller RMSE values in the test data and a smaller difference between the RMSE values of the train and test data indicate better predictive performance.

values in test data with the paper model, which includes the GT indicators, population, and province-level fixed effects. The second-best test data RMSE value is observed for the model that relies only on GT indicators (SPR and SP dummy variables) and fixed effects. The two models that perform the best at predictions on randomly selected data have higher RMSE values in test data than the models that highlight the GT indicators more. Furthermore, these models show a higher difference between the RMSE values in train and test data, indicating more overfitting. It should be noted that the GDP and population variables are measured annually, while the dependent variable to be predicted (SuTP stock) is measured weekly. This may explain the relative underperformance of more traditional indicators, as they would estimate a too-stable pattern, an issue underlined also by Böhme, Gröger and Stöhr (2020).

To better understand the prediction results for 2019 by different models, we focus on ten provinces with the highest and ten provinces with the lowest number of SuTP and visualize the predicted and actual values of SuTP to compare their patterns. We present these plots in Figure 5, and Figure 10 in the Appendix, respectively, and interpret the results based on their RMSE. For both cases, the predicted values by the *comparison model* and the *GT+comparison model* mostly overlap and,

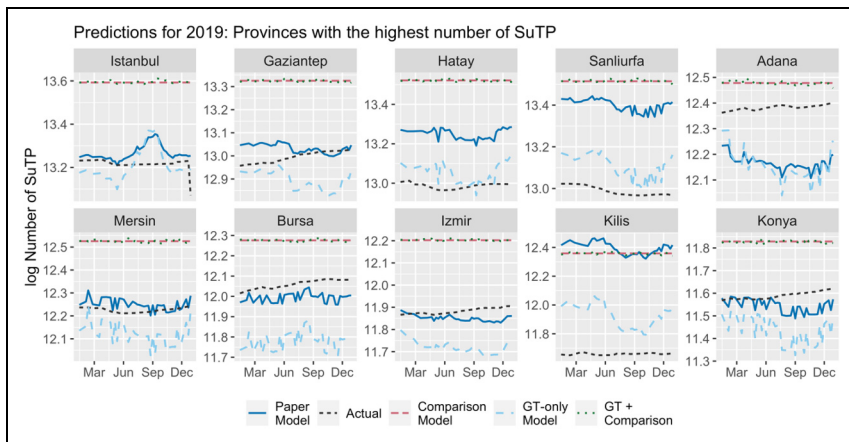


Figure 5. Predictions for the top ten SuTP-hosting provinces — 2019.

Note: Figure 5 shows the prediction results of four different models in comparison to the actual SuTP stocks in the ten provinces that host the highest number of SuTP. The *paper model* predictions, including the online search indicators as well as (log) population and province-level fixed effects, are shown with the straight, blue line. The *comparison model* predictions, including only (log) population and (log) GDP as explanatory variables, are indicated with the red, long-dashed line. The light-blue dash-dotted line shows the *GT-only model* predictions based on the online search indicators. The green dotted line shows *GT + comparison model* predictions. The black short-dashed line shows the actual (log) values of SuTP stocks.

therefore, are hard to distinguish on the graphs. In addition to the stable patterns of SuTP stocks predicted by these two models, they also overestimate the SuTP stock for all ten provinces with the highest number of SuTP.

In the case of ten provinces with the highest number of SuTP, predictions based on the *paper model* are the closest to the actual numbers in six provinces. In three provinces, the *GT-only model* outperforms all others and produces the most accurate predictions. Only in the case of Adana, we see that both the *paper model* and the *GT-only model* underestimate the SuTP stock and the best-performing model for this province is the *GT + comparison model*, despite the overestimation. Of these ten provinces, four are located along the Syrian border (Gaziantep, Hatay, Sanliurfa, and Kilis), two are metropolitan areas relatively close to the border (Adana and Mersin), and four are metropolitan areas relatively far from the border (Istanbul, Bursa, Izmir, and Konya). For all the provinces relatively far from the Syrian border, the *paper model* produces the most accurate predictions, and all three provinces (Hatay, Sanliurfa, and Kilis), for which the *GT-only model* has the most accurate predictions, are located on the border. For all these ten provinces, the best predictions belong to the models that take into account the GT indicators.

In contrast to Figure 5, Figure 10 in the Appendix depicts the predicted values for the ten provinces with the lowest number of SuTP. We highlight two points regarding these prediction results. First, in five¹⁴ of these ten provinces, the *GT-only model* produces the closest values to the actual numbers. Second, in the four provinces with the lowest number of SuTP, an increase predicted by the *paper model* and *GT-only model* corresponds to a decrease in the number of SuTP. In these cases, the models likely misinterpreted an increase in online searches as migration to the province when it was migration out of it. It should be noted that despite this misinterpretation of the pattern in these cases, both models predicted the SuTP stock better than the *comparison model* and the *GT + comparison model*.

Our findings show that including Google Trends predictors in the model increases the predictive power, even if by a small margin, as evidenced in the literature (Choi and Varian 2009). Socio-economic indicators alone (*comparison model*) could be enough for the estimation of randomly selected observations based on reference data from the past and the future and in frequent intervals. However, the *paper model* performs the best when making forecasts on a year at frequent time intervals (such as weekly or monthly) based on data from previous years. In fact, even the model using only Google Trends predictors and province fixed-effects (the *GT-only model*) outperforms the *comparison model* in this case. In line with the findings of Böhme, Gröger and Stöhr (2020), we interpret these results as the advantage of using the finer time detail provided by the Google Trends predictors, compared to the data on socio-economic indicators reported annually and thus forecast a too-stable pattern.

Discussion

In the forced displacement context, register data are limited and often come too late, while survey data collection is challenging due to sensitivity issues and difficulties in the field. Real-time data is much needed for researchers and policy-makers when encountering a sudden flow of refugees, allowing them to better estimate what to expect in the short-term future and plan accordingly. In this study, we shed light on the potential of online searches as a predictor for stocks in the context of forced displacement by examining the case of Syrians under temporary protection in Turkey. Relying on the literature that suggests refugees in general and SuTP in Turkey in specific seek information online using smartphones and the internet (Ulutürk, Uysal and Varol 2019), we show that there is a positive and statistically significant association between the Google Trends Index for province names in Turkey and the number of SuTP across provinces. The association is robust to all the checks and controls applied. Our results corroborate the pattern of a small but significant correlation between online searches and patterns of behavior (Choi and Varian 2009).

¹⁴Giresun, Erzurum, Iğdır, Tunceli and Artvin.

Our findings show that the statistical significance and predictive power of online searches change when time lags are introduced. This result is reasonable considering that searching for information may precede the move and the time needed for preparation before travel. It is also coherent with the necessary bureaucratic steps to be taken, i.e., to register in a new province. Exploring the predictive power of online searches in estimating SuTP stocks, we also find evidence that taking Google Trends variables into account provides better insight into the stock patterns for predictions at short time intervals such as weeks or months. Socioeconomic variables used to predict migrant/refugee stocks, such as population and GDP, are mostly reported annually, and data availability at regional and subregional levels may be problematic. Google Trends data are available at daily, weekly, and monthly intervals depending on the length of the selected time period, and therefore advantageous in forecasting/nowcasting SuTP stock at frequent time intervals. The benefit of using Google Trends in forecasting the pattern of SuTP stocks at more granular detail in time in contrast to the overly stable pattern predicted by the socioeconomic indicators based on a single reporting period corroborates with the findings of Böhme, Gröger and Stöhr (2020) for international migration flows.

Our study contributes to the study of forced displacement using digital trace data. Building on the growing literature on estimating forced migration flows using online trace data (Leasure et al. 2023; Anastasiadou, Volgin and Leasure 2024), we present an example of using online search data to predict refugee stocks in the host country. The case study in our research has certain characteristics that enabled us to use online search data as a predictor. Syrian citizens have arrived in Turkey with the help of the open-border policy and have relative freedom to move inside the country. Surveys suggest a relatively high level of smartphone and internet usage. About 78 to 84% of Syrians in Turkey own a smartphone, and only 10% never use the internet (Sunata 2017; Merisalo and Jauhiainen 2020). Furthermore, differentiation between the local and refugee populations is possible as Turkish and Arabic use different alphabets. While our approach in this study can be applied to other cases of forced displacement, conditions of relatively free movement at the place of study (for stocks) and smartphone/internet usage should be more or less similar. If these conditions are present, our approach could be used to predict refugee accumulation in regions or cities and in almost real-time as the displacement occurs. It may be applied to a single country or the Schengen area in the EU due to free mobility. Future work could also adapt this case study for further research on internally displaced migrants (IDPs) if a differentiation strategy could be established, such as dialects or minority languages.

Although the above-mentioned circumstances enabled the use of online search data as a predictor, we must also acknowledge the limitations. The internet penetration rates of host and refugee populations and the existing digital divides pose a limitation for Google Trends data, as they do for all studies using digital trace data, and may create a bias in forecasting stock patterns. However, despite the challenges of the data and potential bias, we reiterate the argument in the literature that using online

search data offers a useful predictor when data collection is challenging and real-time forecasting is needed for policy decisions. In the context of forced displacement, where data are not easily accessible but very important for researchers, policy-makers, and international organizations online search data offer an alternative source.

When using online search data as a predictor, keyword selection and differentiation of the focus group from the general population are crucial for the research design. The differentiation of refugees from the rest of the online community is generally possible by language and context-related keywords, similar to the strategy of the studies on online search data and international migration. Yet, once the refugees arrive at the destination country, their intentions to move within the country may be harder to track if the destination name is not coupled with the search phrases used since the province names remain unchanged in different languages, save for a few metropolitan areas. The alphabetical difference, if present, is helpful to exploit and isolate the online searches of refugees from the host population, as shown in our study. It must also be acknowledged that while the alphabetical difference helps differentiate between locals and refugees, we cannot claim that the SuTP do all online searches. In our case study, our strategy may capture online searches by Arabic-speaking tourists and unregistered Arabic-speaking migrants along with the online searches of SuTP. While this may create a bias in our results, the positive and significant relationship between SuTP stocks and online searches signal that this bias could be negligible in our study. Future research using online search data in the forced displacement context would be advised to identify a case-specific method to differentiate host and refugee populations. Alphabetical difference, if present, offers an opportunity, yet depending on the keywords of analysis, language differences may suffice. It should be noted that the widespread use of the language of the refugees by other migrants or minorities in the host country may create bias. Such case studies would benefit from accounting for the difficulty of nationality assumption in their research design and aiming to reveal the change in the online search patterns before and after the start of the forced displacement.

Our strategy to use the search popularity ratio (search popularity score of queries made by refugees over that of the locals) offers an extra measure to ensure that the interest in the province is not due to any other incident that may have happened in the said place. This strategy to compare the online searches of locals and migrants, as well as absorb the seasonal interest that may create bias, may also offer insight into future studies on cultural integration of migrants among host populations, where the online language use is the unit of analysis.

This study contributes to the migration literature by analyzing the statistical significance and demonstrating the predictive power of online search popularity in the context of forced displacement and using the case study of the movement of displaced Syrians within Turkey. We further contribute to the growing literature on online search data by exploiting the alphabetical difference to distinguish local and refugee populations and using the search popularity ratio as an explanatory variable. Our study does not pursue a causal relationship between online searches and refugee

stocks but shows an association and uses this association to analyze the predictive power of online search data in the context of forced displacement (Kleinberg et al. 2015). Our findings show that while online search data outperform more commonly used socio-economic variables in predictions at frequent time intervals, online search data are far from outperforming them in random predictions. Based on these results, we conclude that a merged approach would be more beneficial overall, but in cases when real-time and frequent predictions are needed or forced displacement data and/or socio-economic data are lacking at subregional levels, online search data could provide a reliable range. The policy implications of these findings are harder to pin down. Future research may look into developing a set of keywords and phrases that refugees use to search for information online. For better capture of online information-seeking activity, we advise researchers to take alternative spellings and misspellings into account, especially in cases of forced displacement, where people can be assumed to arrive in the country with less information than regular migrants and alphabetical differences may complicate transliteration. Validation of these sets of keywords/phrases would be difficult, but once achieved, coupling these sets of keywords with the names of provinces would provide better information to policy-makers, social workers, and NGOs working with refugees about their needs. Further research on this topic can help urban policy-makers prepare for incoming refugees and national and international policy-makers to plan social assistance programs better.


Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Ebru Sanliturk  <https://orcid.org/0000-0003-4928-8351>

Supplemental Material

Supplemental material for this article is available online.

References

- Anastasiadou, A., A. Volgin, and D. R. Leasure. 2024. "War and Mobility: Using Yandex web Searches to Characterize Intentions to Leave Russia After its Invasion of Ukraine." *Demographic Research* 50: 205–20. <https://doi.org/10.4054/DemRes.2024.50.8>.
- Arora, S. K., Y. Li, J. Youtie, and P. Shapira. 2016. "Using the Wayback Machine to Mine Websites in the Social Sciences: A Methodological Resource." *Journal of the*

- Association for Information Science and Technology* 67 (8): 1904–15. <https://doi.org/10.1002/asi.23503>.
- Askitas, N., and K. F. Zimmermann. 2009. “Google Econometrics and Unemployment Forecasting.” *Applied Economics Quarterly* 55 (2): 107–20. <https://doi.org/10.3790/aeq.55.2.107>.
- Avramescu, A., and A. Wiśniowski. 2021. “Now-casting Romanian Migration into the United Kingdom by Using Google Search Engine Data.” *Demographic Research* 45: 1219–54. <https://www.jstor.org/stable/48640811>.
- Bail, C. A., F. Merhout, and P. Ding. 2018. “Using Internet Search Data to Examine the Relationship Between Anti-Muslim and pro-ISIS Sentiment in US Counties.” *Science Advances* 4 (6): eaao5948. <https://doi.org/10.1126/sciadv.aao5948>.
- Billari, F., F. D’Amuri, and J. Marcucci. 2016. “Forecasting Births Using Google.” In *CARMA 2016–1st International Conference on Advanced Research Methods and Analytics*.
- Böhme, M. H., A. Gröger, and T. Stöhr. 2020. “Searching for a Better Life: Predicting International Migration with Online Search Keywords.” *Journal of Development Economics* 142: 102347. <https://doi.org/10.1016/j.jdeveco.2019.04.002>.
- Brownstein, J. S., C. C. Freifeld, and L. C. Madoff. 2009. “Digital Disease Detection—Harnessing the web for Public Health Surveillance.” *The New England Journal of Medicine* 360 (21): 2153. <https://doi.org/10.1056/NEJMp0900702>.
- Cesare, N. L., T. McCormick, E. Spiro, and E. Zagheni. 2018. “Promises and Pitfalls of Using Digital Traces for Demographic Research.” *Demography* 55 (5): 1979–99. <https://doi.org/10.1007/s13524-018-0715-2>.
- Chang, S.-S., S. S. Kwok, Q. Cheng, P. S. Yip, and Y.-Y. Chen. 2015. “The Association of Trends in Charcoal-Burning Suicide with Google Search and Newspaper Reporting in Taiwan: A Time Series Analysis.” *Social Psychiatry and Psychiatric Epidemiology* 50: 1451–61. <https://doi.org/10.1007/s00127-015-1057-7>.
- Choi, H., and H. Varian. 2009. Predicting Initial Claims for Unemployment Benefits.” *Google Research* (blog), *Google Inc*, July 23. <https://research.google/blog/predicting-initial-claims-for-unemployment-benefits/>
- Chykina, V., and C. Crabtree. 2018. “Using Google Trends to Measure Issue Salience for Hard-to-Survey Populations.” *Socius* 4. <https://doi.org/10.1177/2378023118760414>.
- Connor, P. 2017. The digital footprint of Europe’s refugees. Retrieved June 2022, from Pew Research Center Global Attitudes Project: <https://policycommons.net/artifacts/617838/the-digital-footprint-of-europes-refugees/1598702/>
- De Jong, G. F., B. D. Root, R. W. Gardner, J. T. Fawcett, and R. G. Abad. 1985. “Migration Intentions and Behavior: Decision Making in a Rural Philippine Province.” *Population and Environment* 8 (1): 41–62. <https://doi.org/10.1007/BF01263016>.
- Ettredge, M., J. Gerdes, and G. Karuga. 2005. “Using Web-Based Search Data to Predict Macroeconomic Statistics.” *Communications of the ACM* 48 (11): 87–92. <https://doi.org/10.1145/1096000.1096010>.
- Fantazzini, D., J. Pushchelenko, A. Mironenkov, and A. Kurbatskii. 2021. “Forecasting Internal Migration in Russia Using Google Trends: Evidence from Moscow and Saint Petersburg.” *Forecasting* 3 (4): 774–803. <https://doi.org/10.3390/forecast3040048>.

- Gil-Clavel, S., and E. Zagheni. 2019. "Demographic Differentials in Facebook Usage Around the World." *Proceedings of the International AAAI Conference on Web and Social Media* 13: 647–50. <https://doi.org/10.1609/icwsm.v13i01.3263>.
- Gillespie, M., S. Osseiran, and M. Cheesman. 2018. "Syrian Refugees and the Digital Passage to Europe: Smartphone Infrastructures and Affordances." *Social media + Society* 4 (1). <https://doi.org/10.1177/2056305118764440>.
- Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457 (7232): 1012–4. <https://doi.org/10.1038/nature07634>.
- Google. 2022. *FAQ about Google Trends data*. Retrieved June 2022, from Google Trends: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052
- Guzman, G. 2011. "Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations." *Journal of Economic and Social Measurement* 36 (3): 119–67. <https://doi.org/10.3233/JEM-2011-0342>.
- Hawelka, B., I. Sitko, E. Beinart, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. 2014. "Geo-located Twitter as Proxy for Global Mobility Patterns." *Cartography and Geographic Information Science* 41 (3): 260–71. <https://doi.org/10.1080/15230406.2014.890072>.
- Kirişçi, K., and E. G. Ferris. 2015. "Not Likely to Go Home: Syrian Refugees and the Challenges to Turkey and the International Community." Policy Paper No. 7. The Center on the United States and Europe at Brookings, Turkey Project. <https://www.brookings.edu/articles/not-likely-to-go-home-syrian-refugees-and-the-challenges-to-turkey-and-the-international-community/>
- Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer. 2015. "Prediction Policy Problems." *American Economic Review* 105 (5): 491–5. <https://doi.org/10.1257/aer.p20151023>.
- Leasure, D. R., R. Kashyap, F. Rampazzo, C. A. Dooley, B. Elbers, M. Bondarenko, M. Verhagen, et al. 2023. "Nowcasting Daily Population Displacement in Ukraine Through Social media Advertising Data." *Population and Development Review* 49 (2): 231–54. <https://doi.org/10.1111/padr.12558>.
- Lei, L., and X. Liu. 2022. "The COVID-19 Pandemic and Residential Mobility Intentions in the United States: Evidence from Google Trends Data." *Population, Space and Place* 28 (6): e2581. <https://doi.org/10.1002/psp.2581>.
- Leyesen, B., and P. P. Verhaeghe. 2023. "Searching for Migration: Estimating Japanese Migration to Europe with Google Trends Data." *Quality & Quantity* 57: 4603–31. <https://doi.org/10.1007/s11135-022-01560-0>.
- Lin, A. Y., J. Cranshaw, and S. Counts. 2019. "Forecasting US Domestic Migration Using Internet Search Queries." In *The World Wide Web Conference*, 1061–72. <https://doi.org/10.1145/3308558.3313667>.
- Lorenz, R., J. Beck, S. Horneber, F. Keusch, and C. Antoun. 2022. "Google Trends as a Tool for Public Opinion Research: An Illustration of the Perceived Threats of Immigration." In *Migration Research in a Digitized World*, edited by S. Pöttschke and S. Rinken, 193–206. Cham: Springer.

- McCarthy, M. J. 2010. "Internet Monitoring of Suicide Risk in the Population." *Journal of Affective Disorders* 122 (3): 277–9. <https://doi.org/10.1016/j.jad.2009.08.015>.
- McMahon, S., and N. Sigona. 2021. "Death and Migration: Migrant Journeys and the Governance of Migration During Europe's "Migration Crisis"." *International Migration Review* 55 (2): 605–28. <https://doi.org/10.1177/0197918320958615>.
- Mellon, J. 2014. "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience." *Journal of Elections, Public Opinion & Parties* 24 (1): 45–72. <https://doi.org/10.1080/17457289.2013.846346>.
- Merisalo, M., and J. S. Jauhiainen. 2020. "Digital Divides among Asylum-Related Migrants: Comparing Internet use and Smartphone Ownership." *Tijdschrift Voor Economische en sociale Geografie* 111 (5): 689–704. <https://doi.org/10.1111/tesg.12397>.
- Narli, N. 2018. "Life, Connectivity and Integration of Syrian Refugees in Turkey: Surviving Through a Smartphone." *Questions de Communication* 33: 269–86. <https://doi.org/10.4000/questionsdecommunication.12523>.
- Pelat, C., C. Turbelin, A. Bar-Hen, A. Flahault, and A.-J. Valleron. 2009. "More Diseases Tracked by Using Google Trends." *Emerging Infectious Diseases* 15 (8): 1327. <https://doi.org/10.3201/eid1508.090299>.
- Reis, B. Y., and J. S. Brownstein. 2010. "Measuring the Impact of Health Policies Using Internet Search Patterns: The Case of Abortion." *BMC Public Health* 10 (1): 514. <https://doi.org/10.1186/1471-2458-10-514>.
- Ricketts, C. F., and C. G. Silva. 2017. "An Analysis of Morbidity and Mortality Using Google Trends." *Journal of Human Behavior in the Social Environment* 27 (6): 559–70. <https://doi.org/10.1080/10911359.2017.1299658>.
- Smets, K. 2018. "The Way Syrian Refugees in Turkey Use Media: Understanding "Connected Refugees" Through a Non-media-Centric and Local Approach." *Communications* 43 (1): 113–23. <https://doi.org/10.1515/commun-2017-0041>.
- Solano, P., M. Ustulin, E. Pizzorno, M. Vichi, M. Pompili, G. Serafini, and M. Amore. 2016. "A Google-Based Approach for Monitoring Suicide Risk." *Psychiatry Research* 2016 (246): 581–6. <https://doi.org/10.1016/j.psychres.2016.10.030>.
- Song, T. M., J. Song, J.-Y. An, L. L. Hayman, and J.-M. Woo. 2014. "Psychological and Social Factors Affecting Internet Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends." *Yonsei Medical Journal* 55 (1): 254–63. <https://doi.org/10.3349/ymj.2014.55.1.254>.
- State, B., M. Rodriguez, D. Helbing, and E. Zagheni. 2014. "Migration of Professionals to the US." In *Social Informatics. Lecture Notes in Computer Science*. Vol. 8851, edited by Aiello L. M. and D. McFarland, 531–43. Cham: Springer. https://doi.org/10.1007/978-3-319-13734-6_37.
- Sunata, U. 2017. Syrian Refugee Livelihood Monitor: Summary Assessment. In *Syrian Refugee Livelihood Monitor*, 1–14. Istanbul: INGEV and IPSOS.
- Temporary Protection Regulation, 29153. TC Official Gazette 2014.
- TurkStat. 2023. *Survey on Information and Communication Technology (ICT) Usage Survey in Households and by Individuals*. Ankara: Turkish Statistical Institute.

- Ulutürk, I., I. Uysal, and O. Varol. 2019. Refugee integration in Turkey: a study of mobile phone data for D4R challenge. *Data for refugees challenge workshop*.
- Vicéns-Feliberty, M. A., and C. F. Ricketts. 2016. "An Analysis of Puerto Rican Interest to Migrate to the United States Using Google Trends." *The Journal of Developing Areas* 50 (2): 411–30. <https://www.jstor.org/stable/24737411>.
- Wanner, P. 2021. "How Well can we Estimate Immigration Trends Using Google Data?" *Quality & Quantity* 55 (4): 1181–202. <https://doi.org/10.1007/s11135-020-01047-w>.
- Zagheni, E., and I. Weber. 2015. "Demographic Research with non-Representative Internet Data." *International Journal of Manpower* 36 (1): 13–25. <https://doi.org/10.1108/IJM-12-2014-0261>.
- Wilde, J., W. Chen, and S. Lohmann. 2020. Covid-19 and the future of us fertility: what can we learn from Google? *IZA Discussion Paper*, IZA DP No. 13776.
- Wladyka, D. 2017. "Queries to Google Search as Predictors of Migration Flows from Latin America to Spain." *Journal of Population and Social Studies* 25 (4): 312–27. <https://doi.org/10.25133/JPSSv25n4.002>.
- Yildiz, D., J. Munson, A. Vitali, R. Tinati, and J. A. Holland. 2017. "Using Twitter Data for Demographic Research." *Demographic Research* 37: 1477–514. <https://doi.org/0.4054/DemRes.2017.37.46>.
- Zagheni, E., I. Weber, and K. Gumjadi. 2017. "Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants." *Population and Development Review* 43 (4): 721–34. <https://www.jstor.org/stable/26622775>.

Secondary Sources for the Registries of SuTP

- Arslan Erol, H. 2018. "Göç bağlamında Suriye Türkmen Türkçesi ve Türk dili tarihindeki yeri (Turkish of the Syrian Turkmen Within the Context of Immigration and its Importance in the History of the Turkish Language)." *Akdeniz İnsani Bilimler Dergisi* 8 (2): 33–40. <https://doi.org/10.13114/MJH.2018.408>.
- Erkan, U. 2018. "Ulusçu Asimilasyon/Entegrasyon Politikalarından Medeniyet Perspektifine Dayalı Dirimsel/Diyalojik Etkileşime." In *Göç Sosyolojisi: Farklı Boyutlarıyla Göç (Sociology of Migration: Various Dimensions of Migration)*, edited by R. Şimşek, 49–72. Ankara: Akademisyen Kitabevi.
- Gök, M. 2019. "Kayseri'deki Suriyeli sığınmacıların hayata tutunma stratejileri (Coping mechanisms of Syrian refugees in Kayseri)." Master's thesis, Niğde Ömer Halisdemir Üniversitesi/Sosyal Bilimler Enstitüsü.
- Önder, N. 2019. "Türkiye'de geçici koruma altındaki Suriyelilere yönelik sağlık politikalarının analizi (Analysis of Health Policies for the Syrians Under Temporary Protection in Turkey)." *Göç Araştırmaları Dergisi* 5 (1): 110–65.