Taylor & Francis
Taylor & Francis Group

Check for updates

Forecasting Unemployment Using Internet Search Data via PRISM

Dingdong Yi[a], Shaoyang Ning[b], Chia-Jung Chang[c], and S. C. Kou[d]

[a]Citadel Americas LLC, Chicago, IL; [b]Department of Mathematics and Statistics, Williams College, Williamstown, MA; [c]Department of Statistics and Applied Probability, National University of Singapore, Singapore; [d]Department of Statistics, Harvard University, Cambridge, MA

ABSTRACT

Big data generated from the Internet offer great potential for predictive analysis. Here we focus on using online users' Internet search data to forecast unemployment initial claims weeks into the future, which provides timely insights into the direction of the economy. To this end, we present a novel method Penalized Regression with Inferred Seasonality Module (PRISM), which uses publicly available online search data from Google. PRISM is a semiparametric method, motivated by a general state-space formulation, and employs nonparametric seasonal decomposition and penalized regression. For forecasting unemployment initial claims, PRISM outperforms all previously available methods, including forecasting during the 2008–2009 financial crisis period and near-future forecasting during the COVID-19 pandemic period, when unemployment initial claims both rose rapidly. The timely and accurate unemployment forecasts by PRISM could aid government agencies and financial institutions to assess the economic trend and make well-informed decisions, especially in the face of economic turbulence.

1. Introduction

Driven by the growth and wide availability of Internet and online platforms, big data are generated with an unprecedented speed nowadays. They offer the potential to inform and transform decision making in industry, business, social policy and public health (Manyika et al. 2011; Chen, Chiang, and Storey 2012; McAfee and Brynjolfsson 2012; Murdoch and Detsky 2013; Khoury and Ioannidis 2014; Kim, Trimi, and Chung 2014). Big data can be used for developing predictive models for systems that would have been challenging to predict with traditional small-sample-based approaches (Siegel 2013; Einav and Levin 2014). For instance, numerous studies have demonstrated the potential of using Internet search data in tracking influenza outbreaks (Ginsberg et al. 2009; Yang, Santillana, and Kou 2015; Ning, Yang, and Kou 2019; Yang, Ning, and Kou 2021), dengue fever outbreaks (Yang et al. 2017), financial market returns (Preis, Moat, and Stanley 2013; Risteski and Davcev 2014), consumer behaviors (Goel et al. 2010), unemployment (Ettredge, Gerdes, and Karuga 2005; Choi and Varian 2012; Li 2016), and housing prices (Wu and Brynjolfsson 2015).

We consider here using Internet users' Google search to forecast US unemployment initial claims weeks into the future. Unemployment initial claims measure the number of jobless claims filed by individuals seeking to receive state jobless benefits. It is closely watched by the government and the financial market, as it provides timely insights into the direction of the economy. A sustained increase of initial claims would indicate rising unemployment and a challenging economy, whereas a steady decrease of initial claims would signal recovery of labor market. During the great financial crisis of

2008 and the COVID-19 pandemic, these unemployment data have been a key focus for government agencies when making fiscal and monetary policy decisions under unprecedented pressure.

Weekly unemployment initial claim is the (unadjusted) total number of actual initial claims filed under the Federal-State Unemployment Insurance Program in each week ending on Saturday. The Employment and Training Administration (ETA) of the U.S. Department of Labor collects the weekly unemployment insurance claims reported by each state's unemployment insurance program office, and releases the data to the public at 8:30 a.m. (eastern time) on the following Thursday. Thus, the weekly unemployment initial claim data are reported with a one-week delay: the number reported on Thursday of a given week is actually the unemployment initial claim number of the preceding week. For accessing the general economic trend, it is, therefore, highly desirable for government agencies and financial institutions to predict the unemployment situation of the current week, which is known as nowcasting (Giannone, Reichlin, and Small 2008), as well as weeks into the future. In this article, we use the general phrase *forecasting* to cover both nowcasting (the current week) and predicting into future weeks.

In contrast to the one-week delayed unemployment reports by the Department of Labor, Internet users' online search of unemployment-related query terms provides highly informative and *real-time* information for the current unemployment situation. For instance, a surge of Internet search of "unemployment office," "unemployment benefits," "unemployment extension," etc. in a given week could indicate an increase of unemployment of *that* week, as presumably more people unemployed
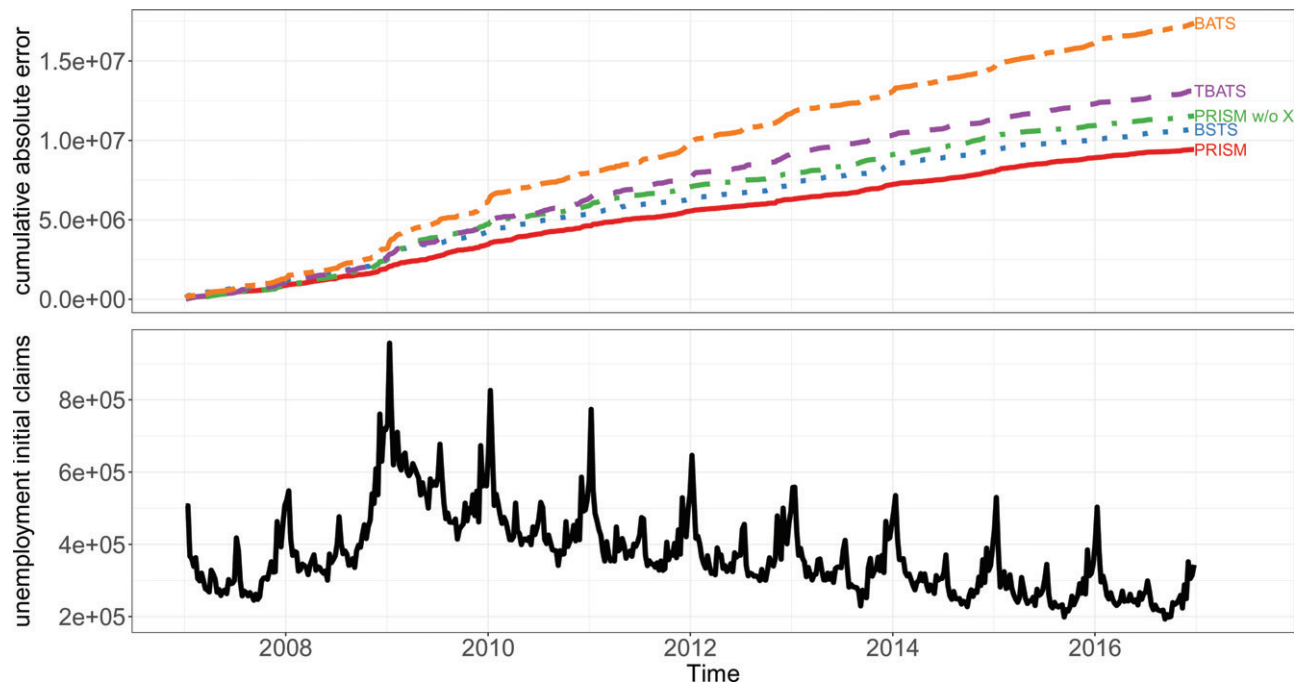
Figure 1. (Top) The cumulative absolute error of nowcasting of different methods. (Bottom) The unemployment initial claims for the same period of 2007–2016.

are searching for information of getting unemployment aid. Internet search data, offering a real-time "peek" of the current week, thus, augment the delayed official time-series unemployment data.

There are several challenges in developing an effective method to forecast weekly unemployment initial claims with Internet search data. First, the volatile seasonality pattern accounts for most of the variation of the target time series. Figure 1 (bottom) plots the weekly unemployment initial claims from 2007 to 2016; the seasonal spikes are particularly noteworthy. A prediction method should address and use the strong seasonality to achieve good prediction performance. Second, the method needs to effectively incorporate the most up-to-date Internet search data into the modeling of target time series. Third, as people's search pattern and the search engine both evolve over time, the method should be able to accommodate this dynamic change.

Most time series models rely on state-space models to deal with seasonality, where the latent components capture the trend and seasonality (Aoki 1987; Harvey 1989). Among the time series models, structural time series models and innovation state-space models are two main frameworks (Harvey and Koopman 1993; Hyndman et al. 2008; Durbin and Koopman 2012), both of which have various extensions of seasonal pattern modeling and can incorporate exogenous signals as regression components (see Supplementary Material A2 for more discussion). For nowcasting time series with seasonal pattern, Scott and Varian (2013, 2014) developed a Bayesian method based on the structural time series model, using a spike-and-slab prior for variable selection, and applied it to nowcast unemployment initial claims with Google search data by treating the search data as regressors. Alternative to this regression formulation, Banbura et al. (2013) proposed a nowcasting method using a

factor model, in which target time series and related exogenous time series are driven by common factors.

Here we introduce a novel prediction method PRISM, which stands for Penalized Regression with Inferred Seasonality Module, for forecasting times series with seasonality, and use it to forecast unemployment initial claims. Our method is semiparametric in nature, and takes advantage of both the state-space formulation for time series forecasting and penalized regression. With the semiparametric method PRISM, we significantly expand the range of time series models for forecasting, going beyond the traditional approaches, which are often tailored for individual parametric models. PRISM offers a robust and more accurate forecasting alternative to traditional parametric approaches and effectively addresses the three aforementioned challenges in forecasting time series with strong seasonality. First, our method accommodates various nonparametric and model-based seasonal decomposition tools, and effectively incorporates the estimated seasonal components into predictive modeling. It thus can robustly handle complex seasonal patterns. Second, different from the traditional regression formulation, our joint modeling of the target time series and the exogenous variables accommodates the potential causal relationship between them—people do online Google search in response of becoming unemployed. Third, PRISM uses dynamic forecasting—training its predictive equation each week for the forecasting—and uses rolling window and exponential weighting to account for the time-varying relationship between the target time series and the exogenous variables. For forecasting unemployment initial claims, PRISM delivers superior performance over all existing forecasting methods for the entire time period of 2007–2019, and is exceptionally robust to the ups and downs of the general economic environment, including the huge volatility caused by the 2008 financial crisis.

While the forecasting target here is the unemployment initial claims, we want to highlight that PRISM applies to forecasting other time series with complex seasonal patterns.

2. Data and Method

2.1. Initial Claims Data and Internet Search Data From Google

The weekly (nonseasonally adjusted) initial claims are our target time series. The initial claims for the preceding week are released every Thursday. The time series of the initial claims from 1967 to present are available at *https://fred.stlouisfed.org/series/ICNSA*.

The real-time Internet search data we used were obtained from Google Trends (*www.google.com/trends*) with Python package `pytrends`. The Google Trends website, which is publicly accessible, provides weekly (relative) search volume of search query terms specified by a user. Specifically, for a user-specified query term, Google Trends provides integer-valued weekly times series (after 2004); each number in the time series, ranging from 0 to 100, represents the search volume of that search query term in a given week divided by the total online search volume of that week; and the number is normalized to integer values from 0 to 100, where 100 corresponds to the maximum weekly search within the time period (specified by the user). Figure A1 in the supplementary materials illustrates the Google Trend time series of several search query terms in a 5-year span.

The search query terms in our study were also identified from the Google Trends tool. One feature of Google Trends is that, in addition to the time series of a specific term (or a general topic), it also returns the top query terms that are most highly correlated with the specific term. In our study, we used a list of top 25 Google search terms that are most highly correlated with the term "unemployment." Table A1 of the supplementary materials lists these 25 terms, which were generated by Google Trends on January 11, 2018; they included 12 general unemployment related query terms, such as "unemployment office," "unemployment benefits" and "unemployment extension," as well as 13 query terms that were combinations of state names and "unemployment," such as "California unemployment" and "unemployment Florida."

2.2. Overview of PRISM

PRISM employs a two-stage estimation procedure for forecasting time series y_t using its lagged values and the available exogenous variables x_t. The derivation and rationale of each step will be described subsequently.

Input: Target time series $\{y_{1:(t-1)}\}$ and exogenous time series $\{x_{t_0:t}\}$. In the forecasting of unemployment initial claims, $\{y_{1:(t-1)}\}$ is the official weekly unemployment initial claim data reported with one-week delay, and $\{x_{t_0:t}\}$ is the multivariate Google Trends data starting from 2004.

Stage 1 of PRISM: Seasonal decomposition. Decompose $\{y_t\}$, the univariate time series of interest, into the seasonal component $\{\gamma_t\}$, and the seasonally adjusted component $z_t = y_t - \gamma_t$. In particular, with a fixed rolling window length

M, Stage 1 nonparametrically decomposes $\{y_{(t-M):(t-1)}\}$ into estimated seasonal component $\{\hat{\gamma}_{i,t}\}_{i=(t-M),...,(t-1)}$ and estimated seasonally adjusted component $\{\hat{z}_{i,t}\}_{i=(t-M),...,(t-1)}$ using data available at time t.

Stage 2 of PRISM: Penalized regression. Forecast target time series using:

$$\hat{y}_{t+l} = \mu_y^{(l)} + \sum_{j=1}^{K} \alpha_j^{(l)} \hat{z}_{t-j,t} + \sum_{j=1}^{K} \delta_j^{(l)} \hat{\gamma}_{t-j,t} + \sum_{i=1}^{P} \beta_i^{(l)} x_{i,t}, \quad (1)$$

where the coefficients are estimated by a rolling-window L_1 penalized linear regression using historical data for each forecasting horizon: $l = 0$ corresponds to nowcast; $l \geq 1$ corresponds to forecasting future weeks. Note that for notational ease, we will suppress "(l)" in the superscripts of the coefficients in the subsequent discussion.

2.3. Derivation of PRISM

PRISM is motivated by a general state-space formulation for univariate time series with seasonal pattern. We postulate that the seasonal and seasonally adjusted component $\{\gamma_t\}$ and $\{z_t\}$ each evolve according to a linear state-space model with state vectors $\{s_t\}$ and $\{h_t\}$, respectively:

$$y_t = z_t + \gamma_t, \quad (2a)$$

$$\begin{cases} z_t = w'h_t + \epsilon_t, & (2b) \\ h_t = Fh_{t-1} + \eta_t, & (2c) \end{cases} \quad \begin{cases} \gamma_t = v's_t + \zeta_t, & (2d) \\ s_t = Ps_{t-1} + \omega_t, & (2e) \end{cases}$$

where $(\epsilon_t, \zeta_t, \eta_t', \omega_t')' \overset{iid}{\sim} \mathcal{N}(0, H)$, and $\theta = (w, F, v, P, H)$ are the parameters.

Our state-space formulation contains a variety of widely used time series models, including structural time series models (Harvey 1989) and additive innovation state-space models (Aoki 1987; Ord, Koehler, and Snyder 1997; Hyndman et al. 2008). Under the general formulation (2), a specific parametric model can be obtained by specifying the state vectors $\{h_t\}$ and $\{s_t\}$ along with the dependence structure H. We highlight a few special cases of model (2) in the supplementary materials.

PRISM also models the contemporaneous information from exogenous variables. Let $x_t = (x_{1,t}, x_{2,t}, \ldots, x_{p,t})'$ be the vector of the exogenous variables at time t. We postulate a state-space model for x_t on top of y_t, instead of adding them as regressors as in traditional models. In particular, at each time t, we assume a multivariate normal distribution for x_t conditional on the level of unemployment initial claims y_t,

$$x_t|y_t \sim \mathcal{N}_p(\mu_x + y_t\beta, Q), \quad (3)$$

where $\beta = (\beta_1, \ldots, \beta_p)'$, $\mu_x = (\mu_{x_1}, \ldots, \mu_{x_p})'$, and Q is the covariance matrix. x_t is assumed to be independent of $\{y_l, x_l : l < t\}$ conditional on y_t. For $\{y_t\}$ following the general state-space model (2), our joint model for y_t and x_t can be diagrammed as:

$$\begin{array}{ccccc} \cdots & \to & (s_t, h_t) & \to & (s_{t+1}, h_{t+1}) & \to & \cdots \\ & & \downarrow & & \downarrow & & \\ & & y_t & & y_{t+1} & & \\ & & \downarrow & & \downarrow & & \\ & & x_t & & x_{t+1} & & \end{array}$$
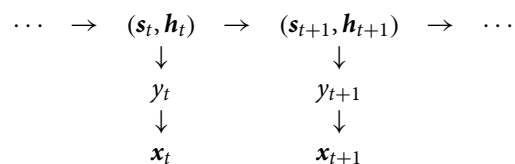
To forecast y_{t+l} under above model assumptions at time t, we consider the predictive distribution of y_{t+l} by conditioning on the historical data $\{y_{1:(t-1)}\}$ and contemporaneous exogenous time series $\{\boldsymbol{x}_{t_0:t}\}$ as well as the latent seasonal component $\{\gamma_{1:(t-1)}\}$. $z_{1:(t-1)}$ is known given $y_{1:(t-1)}$ and $\gamma_{1:(t-1)}$. We can derive a universal representation of the predictive distribution $p(y_{t+l}|z_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t})$, which is normal with mean linear in $z_{1:(t-1)}$, $\gamma_{1:(t-1)}$ and $\boldsymbol{x}_t$ as in (1) (see the supplementary materials for the proof). This observation leads to our two-stage semiparametric estimation procedure PRISM for nowcasting y_t and forecasting y_{t+l} ($l \geq 1$) using all available information at time t.

2.4. Stage 1 of PRISM: Seasonal Decomposition

PRISM estimates the unobserved seasonal components $\gamma_{1:(t-1)}$ in the first stage. For this purpose, various seasonal decomposition methods can be used here, including nonparametric methods such as the classical additive seasonal decomposition (Stuart and Kendall 1963) and parametric methods based on innovation state-space models. We use the method of Seasonal and Trend decomposition using Loess (STL) (Cleveland et al. 1990) as the default choice. The STL method is nonparametric. It is widely used and robust for decomposing time series with few assumptions owing to its nonparametric nature.

At every time t for forecasting, we apply the seasonal decomposition method (such as the default STL) to historical initial claims observations $y_{(t-M):(t-1)}$ with M being a large number. For each rolling window from $t - M$ to $t - 1$, the univariate time series $y_{(t-M):(t-1)}$ is decomposed into three components: seasonal, trend, and the remainder. Denote $\hat{\gamma}_{i,t}$ and $\hat{z}_{i,t}$ as the estimates of γ_i and z_i using data available at time t. Then, at each t the seasonal decomposition generates estimated seasonal component time series $\{\hat{\gamma}_{i,t}\}_{i=(t-M),...,(t-1)}$ and seasonally adjusted time series $\{\hat{z}_{i,t}\}_{i=(t-M),...,(t-1)}$; the latter is the sum of trend component and remainder component. In our forecasting of unemployment initial claims, we took $M = 700$. We describe the basic procedure of the default STL in the supplementary materials.

2.5. Stage 2 of PRISM: Penalized Linear Regression

For each fixed forecasting horizon l (≥ 0), we estimate y_{t+l} by the linear predictive equation:

$$\hat{y}_{t+l} = \mu_y + \sum_{j=1}^{K} \alpha_j \hat{z}_{t-j,t} + \sum_{j=1}^{K} \delta_j \hat{\gamma}_{t-j,t} + \sum_{i=1}^{p} \beta_i x_{i,t}, \quad (4)$$

where for notational ease we have suppressed l in the coefficients and used the generic notations $\mu_y, \alpha_j, \delta_j$ with the understanding that there is a separate set of $\{\mu_y, \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K), \boldsymbol{\delta} = (\delta_1, \ldots, \delta_K), \boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)\}$ for each l. At each time t and for each forecasting horizon l, the regression coefficients $\mu_y, \boldsymbol{\alpha}, \boldsymbol{\delta}$

and $\boldsymbol{\beta}$ are obtained by minimizing

$$\frac{1}{N} \sum_{\tau=t-l-N}^{t-l-1} w^{t-\tau} \Big(y_{\tau+l} - \mu_y - \sum_{j=1}^{K} \alpha_j \hat{z}_{\tau-j,\tau} - \sum_{j=1}^{K} \delta_j \hat{\gamma}_{\tau-j,\tau}$$
$$- \sum_{i=1}^{p} \beta_i x_{i,\tau} \Big)^2 + \lambda_1 \left(||\boldsymbol{\alpha}||_1 + ||\boldsymbol{\delta}||_1 \right) + \lambda_2 ||\boldsymbol{\beta}||_1, \tag{5}$$

where N is the length of a rolling window, w is a discount factor, and λ_1 and λ_2 are nonnegative regularization parameters.

2.6. Features of PRISM

There are several distinct features of our estimation procedure. First, a rolling window of length N is employed. This is to address the fact that the parameters in the predictive equation (4) can vary with time t. In our case, people's search pattern and the search engine tend to evolve over time, and it is quite likely that the same search phrases would contribute in different ways over time to the response variable. Correspondingly, the coefficients in (4) need to be estimated dynamically each week, and the more recent observations should be considered more relevant than the distant historical observations for inferring the predictive equations of current time. The rolling window of observations and the exponentially decreasing weights are utilized for such purpose. Our use of exponential weighting is related to the weighted least square formulation that is usually referred to as discounted weighted regression in the econometrics literature (Ameen and Harrison 1984; Taylor 2010).

Second, since the number of unknown coefficient in (4) tends to be quite large compared to the number of observations within the rolling window, we applied L_1 regularization in our rolling-window estimation (Tibshirani 1996), which gives robust and sparse estimate of the coefficients. Up to two L_1 penalties are applied: on $(\boldsymbol{\alpha}, \boldsymbol{\delta})$ and on $\boldsymbol{\beta}$, as they represent two sources of information: information from time series components $\{\hat{z}_t\}$ and $\{\hat{\gamma}_t\}$, and information from the exogenous variables $\{\boldsymbol{x}_t\}$.

Third, PRISM is a semiparametric method. The predictive equation (4) is motivated and derived from our state-space formulation (2). However, the estimation is not parametric in that (i) the seasonal and seasonally adjusted components are learned nonparametrically in Stage 1, and (ii) the coefficients in (4) are dynamically estimated each week in Stage 2. Combined together, the two stages of PRISM give us a simple and robust estimation procedure. This approach is novel and different from the typical approaches for linear state-space models, which often estimate unknown parameters using specific parameterization and select a model based on information criteria (Hyndman et al. 2008).

2.7. Using PRISM Without Exogenous Variables

In the case when exogenous time series $\{\boldsymbol{x}_t\}$ are not available, PRISM estimates y_{t+l} according to the following linear predictive equation:

$$\hat{y}_{t+l} = \mu_y + \sum_{j=1}^{K} \alpha_j \hat{z}_{t-j,t} + \sum_{j=1}^{K} \delta_j \hat{\gamma}_{t-j,t}, \tag{6}$$

which is a degenerated special case of the predictive equation (4). Under the same estimation procedure as in (5) except that $\boldsymbol{\beta}$ and $\boldsymbol{x}_t$ are dropped, predictive Equation (6) can be used to forecast univariate time series with seasonal patterns without exogenous time series.

2.8. Constructing Point-Wise Predictive Intervals for PRISM Estimate

The semiparametric nature of PRISM makes it more difficult to construct predictive intervals on PRISM forecasts, as we cannot rely on parametric specifications, such as posterior distributions, for predictive interval construction. However, the fact that we are forecasting time series suggests a (nonparametric) method for us to construct predictive intervals based on the historical performances of PRISM.

For nowcasting at time t, given the historical data available up to time $t - 1$, we can evaluate the root mean square error of nowcasting for the last L time periods as

$$\hat{se}_t = \left(\frac{1}{L} \sum_{\tau=t-L}^{t-1} (\hat{y}_\tau - y_\tau)^2\right)^{1/2},$$

where $\hat{y}_\tau$ is the real time PRISM estimate for y_τ generated at time τ. Under the assumption of local stationarity and normality of the residual, $\hat{se}_t$ would be an estimate for the standard error of $\hat{y}_t$. We can thus use it to construct predictive interval for the current PRISM estimate. An $1 - \alpha$ point-wise predictive interval is given by $(\hat{y}_t - z_{\alpha/2}\,\hat{se}_t,\ \hat{y}_t + z_{\alpha/2}\,\hat{se}_t)$, where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. Supplementary Material A10 shows that the empirical residuals are approximately normal, supporting our construction. The point-wise predictive intervals for forecasting into future weeks can be constructed similarly. In practice, we set $L = 52$, estimating the $\hat{se}_t$ based on the forecasts of the most recent one-year window.

2.9. Training PRISM for Forecasting Unemployment Initial Claims

In our forecasting of unemployment initial claims, we applied a 3-year rolling window of historical data to estimate the parameters in (5), that is, $N = 156$ (weeks). The choice of 3-year rolling window is recommended in the literature (D'Amuri and Marcucci 2017) as well as supported by our empirical studies (Supplementary A7). In addition, since the Google Trends data are only available since 2004, with the 3-year rolling window we are able to test the performance of PRISM in 2007–2009, the entire span of the financial crisis, which serves as an important test of the capability of the various prediction methods. We took $K = 52$ (weeks) to employ the most recent 1-year estimated seasonal and seasonally adjusted components, and $p = 25$ (Google search terms) according to the list of top 25 nationwide query terms related to "unemployment" in Table A1.

The weekly search volume data from Google Trends are limited up to a 5-year span per download. The subsequent normalization of the search volumes done by Google is based on the search query term and the specified date range, which implies that the absolute values of search volumes are normalized differently between different 5-year ranges. However, within the same set of 5-year data the relative scale of variables is consistent (as they are normalized by the same constant). Therefore, to avoid the variability from different normalization across different 5-year spans, and to ensure the coherence of the model, for each week, we used the same set of 5-year-span data downloaded for both the training data and the prediction.

For the choice of the discount factor, we took $w = 0.985$ as the default choice. This follows the suggestion by Lindoff (1997) that setting the discount factor between 0.95 and 0.995 works in most applications. We further conducted the experiments for $w \in [0.95, 0.995]$ (see Supplementary Material A6) and found the performance of PRISM is quite robust for $w \in [0.95, 0.995]$ while $w = 0.985$ gives optimal accuracy for our in-sample nowcasting.

For the regularization parameters λ_1 and λ_2 in (5), we used cross-validation to minimize the mean squared predictive errors (i.e., the average of prediction errors from each validation set of data). We found empirically that the extra flexibility of having two separate λ_1 and λ_2 does not give improvement over fixing $\lambda_1 = \lambda_2$. In particular, we found that for every forecasting horizon $l = 0, 1, 2, 3$, in the cross-validation process of setting (λ_1, λ_2) for separate L_1 penalty, over 80% of the weeks showed that the smallest cross-validation mean squared error when restricting $\lambda_1 = \lambda_2$ is within 1 standard error of the global smallest cross-validation mean squared error. For model simplicity, we thus chose to further restrict $\lambda_1 = \lambda_2$ when forecasting unemployment initial claims.

2.10. Accuracy Metrics

We used root-mean-squared error (RMSE) and mean absolute error (MAE) to evaluate the performance of different methods. For an estimator $\{\hat{y}_t\}$ and horizon l, the RMSE and MAE are defined, respectively, as $\text{RMSE}_l(\hat{\boldsymbol{y}}, \boldsymbol{y}) = [\frac{1}{(n_2-n_1-l+1)} \sum_{t=n_1+l}^{n_2} (\hat{y}_t - y_t)^2]^{1/2}$ and $\text{MAE}_l(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{(n_2-n_1-l+1)} \sum_{t=n_1+l}^{n_2} |\hat{y}_t - y_t|$, where $n_1 + l$ and n_2 are, respectively, the start and end of the forecasting period for each l.

3. Results

3.1. Retrospective Forecasting for 2007–2016

We applied PRISM to produce forecasts of weekly unemployment initial claims for the time period of 2007 to 2016 for four time horizons: real-time, one, two, and three weeks ahead of the current time. We compared the forecasts to the ground truth—the unemployment initial claims released by the Department of Labor one-week behind real-time—by measuring the RMSE and MAE.

For comparison, we calculated the RMSE and MAE of four alternative forecasting methods: (a) Bayesian structural time series (BSTS) (Scott and Varian 2014); (b) and (c), two forecasting methods using exponential smoothing: BATS and TBATS (De Livera, Hyndman, and Snyder 2011); and (d) the naive method, which without any modeling effort simply uses the last available weekly unemployment initial claims number (which is of the prior week) as the prediction for the current week,

one, two, and three week(s) later. The naive method serves as a baseline. Both BATS and TBATS are based on innovation state-space model; BATS is an acronym for key features of the model: Box–Cox transformation, ARMA errors, trend, and seasonal components; TBATS extends BATS to handle complex seasonal patterns with trigonometric representations, and the initial T connotes "trigonometric." BSTS only produces nowcasting; it does not produce numbers for forecasting into future weeks.

As PRISM uses both historical unemployment initial claims data and Google search information, to quantify the contribution of the two resources, we also applied PRISM but without the Google search information. We denoted this method as "PRISM w/o x_t."

For fair comparison, in generating retrospective estimates of unemployment initial claims, we reran all methods each week using only the information available up to that week, that is, we obtained the retrospective estimation as if we had relived the testing period of 2007–2016. The two exponential smoothing methods (b) BATS and (c) TBATS only use historical initial claims data and do not offer the option of including exogenous variable in their forecasting, while the method (a) BSTS allows the inclusion of exogenous variables. Thus, for forecasting at each week t, BSTS takes the historical initial claim data and Google Trends data as input, whereas BATS and TBATS use historical initial claim data only. PRISM was applied twice: with and without Google search information. The results of BSTS, BATS and TBATS were produced by their respective R packages under their default settings.

Table 1 presents the performance of forecasting (including nowcasting) unemployment initial claims over the entire period of 2007–2016 for the four forecasting horizons. The RMSE and MAE numbers reported here are relative to the naive method, that is, the number reported in each cell is the ratio of the error of a given method to that of the naive method. The absolute error of the naive method is reported in the parentheses. BSTS does not produce numbers for forecasting into future weeks, as its R package outputs prediction of the target time series only with exogenous variables inputted.

Table 1. Performance of different methods over 2007–2016 for four forecasting horizons: real-time, 1 week, 2 weeks, and 3 weeks ahead.

	Real-time	1 week	2 weeks	3 weeks
RMSE				
PRISM	**0.493**	**0.483**	**0.461**	**0.470**
PRISM w/o x_t	0.647	0.532	0.507	0.524
BSTS	0.588	–	–	–
BATS	1.002	0.897	0.848	0.832
TBATS	0.711	0.559	0.544	0.528
naive	1 (50551)	1 (62227)	1 (69747)	1 (73527)
MAE				
PRISM	**0.539**	**0.517**	**0.476**	**0.460**
PRISM w/o x_t	0.659	0.559	0.510	0.496
BSTS	0.612	–	–	–
BATS	0.992	0.898	0.825	0.781
TBATS	0.750	0.599	0.570	0.525
naive	1 (33637)	1 (41121)	1 (47902)	1 (52794)

NOTE: RMSE and MAE here are relative to the error of naive method; that is, the number reported is the ratio of the error of a given method to that of the naive method; the absolute RMSE and MAE of the naive method are reported in the parentheses. The boldface indicates the best performer for each forecasting horizon and each accuracy metric.

Table 1 reveals the following. First, PRISM uniformly outperforms all the other methods for the entire period of 2007–2016 under all forecasting horizons. Second, the real-time Google Trends data are very helpful for nowcasting, as PRISM and BSTS have better nowcasting results than the other methods that use only historical initial claim data. Third, the contribution of contemporaneous Google Trends data becomes less significant in forecasting future weeks, as evidenced by the shrinking performance gap between PRISM and "PRISM w/o x_t" from nowcasting to forecasting. Fourth, among the three methods that only use historical initial claim data, the predictive method based on PRISM without Google information outperforms the exponential smoothing methods BATS and TBATS.

Following the suggestion of a referee, we further compared the performance of PRISM to the other methods with an additional metric: cumulative sum of squared forecast error differences (CSSED) (Welch and Goyal 2008), which calculates the cumulative difference in mean-squared error (MSE) between PRISM and the alternative. The CSSED at time T for an alternative method m is defined as $\text{CSSED}_{m,\text{PRISM}} = \sum_{t=1}^{T}(e_{t,m}^2 - e_{t,\text{PRISM}}^2)$, where $e_{t,m}$ and $e_{t,\text{PRISM}}$ are the prediction errors at time t for method m and PRISM, respectively. The detailed comparison results are given in Supplementary Material A13, which again shows that the advantage of PRISM over alternatives is consistent over the whole evaluation period.

To assess the statistical significance of the improved prediction power of PRISM compared to the alternatives, we conducted Diebold–Mariano test (Diebold and Mariano 1995), which is a nonparametric test for comparing the prediction accuracy between two time-series forecasting methods. Table 2 reports the p-values of the Diebold–Mariano test (the null hypothesis being that PRISM and the alternative method in comparison have the same prediction accuracy in RMSE). With all the p-values smaller than 2.1%, Table 2 shows that the improved prediction accuracy of PRISM over BSTS, BATS, and TBATS is statistically significant in all of the forecasting horizons evaluated. Further comparison with two additional methods is presented in Supplementary Material A12, where PRISM continues to show significant advantage in prediction accuracy over seasonal AR model and the method of D'Amuri and Marcucci (2017).

Figure 2 shows the RMSE of the yearly nowcasting results of the different methods; here the RMSE is measured relative to the error of the naive method. It is seen that PRISM gives consistent relative RMSE throughout the 2007–2016 period. It is noteworthy that PRISM outperforms all other methods in 2008 and 2009 when the financial crisis caused significant instability

Table 2. p-values of the Diebold–Mariano test for prediction accuracy comparison between PRISM and the alternatives over 2007–2016 for four forecasting horizons: real-time, 1 week, 2 weeks, and 3 weeks ahead.

	Real-time	1 week	2 weeks	3 weeks
PRISM w/o	7.86×10^{-5}	2.09×10^{-2}	2.10×10^{-2}	3.95×10^{-3}
BSTS	7.14×10^{-3}	–	–	–
BATS	9.17×10^{-8}	3.60×10^{-8}	1.05×10^{-9}	1.88×10^{-9}
TBATS	5.20×10^{-9}	7.24×10^{-3}	2.59×10^{-3}	1.86×10^{-2}

NOTE: The null hypothesis of the test is that PRISM and the alternative method in comparison have the same prediction accuracy in RMSE.
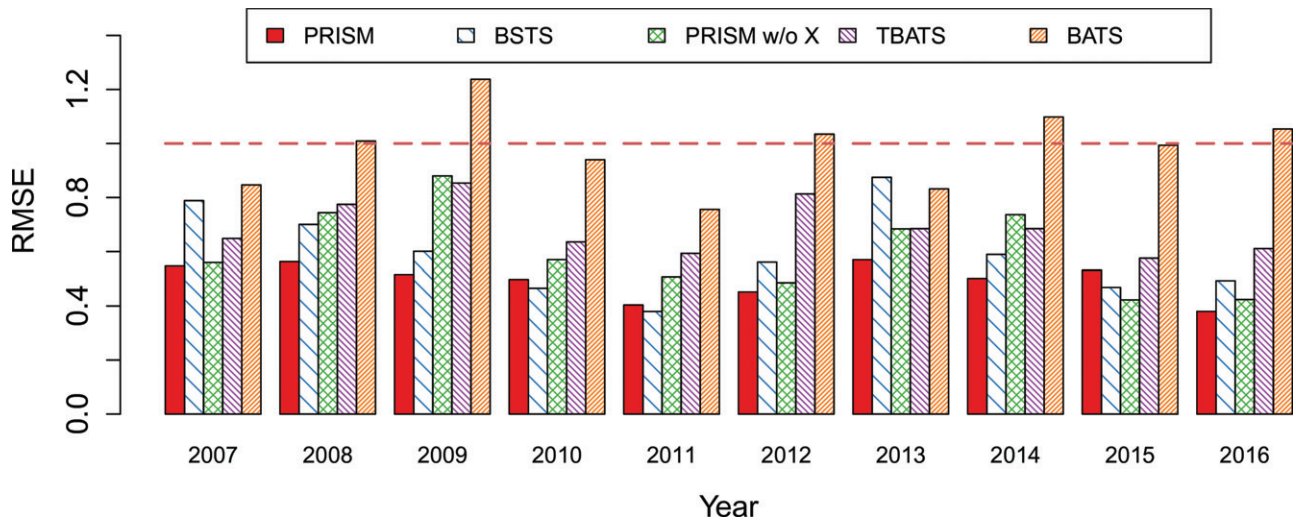
Figure 2. Yearly nowcasting performance of different methods from 2007 to 2016. RMSE is measured relative to the error of the naive method; a value above 1 indicates that the method performs worse than the naive method in that time period.
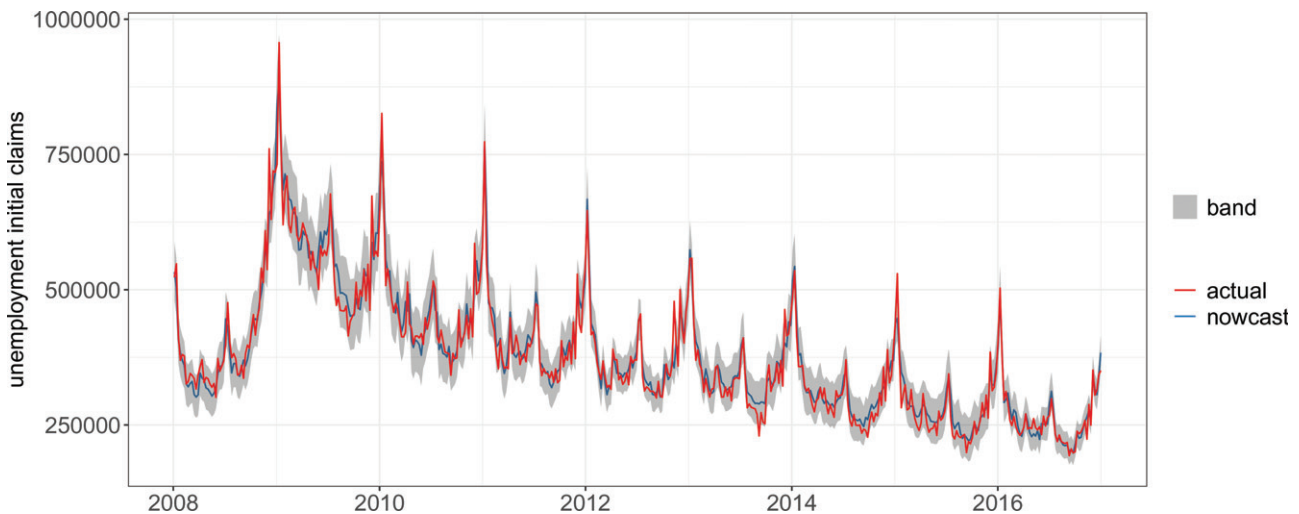

Figure 3. Predictive Interval of PRISM from 2008 to 2016. The shaded area corresponds to the 95% point-wise predictive interval of PRISM nowcasting. The blue curve is the point estimate of PRISM nowcasting. The red curve is the true unemployment initial claims. The actual coverage of the 95% PRISM predictive interval is 96.6% in 2008–2016.

in the U.S. economy. For predictions into future weeks, PRISM also gives the leading performance for the 2007–2016 period (detailed year-by-year plots for the forecasting performance of the different methods for each forecasting horizon are provided in Supplementary Material A14).

For a closer look of the performance of different methods, Figure 1 (top) shows how the absolute errors of nowcasting accumulate through 2007 to 2016. The cumulative absolute error of PRISM rises at the slowest rate among all methods. As shown in Figure 1 (bottom), the 2008 financial crisis caused significantly more unemployment initial claims. PRISM handles the financial crisis period well, as the accumulation of error is rather smooth for the financial crisis period. Other methods all accumulate loss in a considerably higher rate during the financial crisis. Furthermore, PRISM handles the strong seasonality of initial claim data well, since the accumulation of error is smooth within each year. Among all the methods considered, BATS is bumpy in handling seasonality, as the accumulation jumps when the initial claim data spikes.

We further constructed point-wise predictive intervals for the PRISM estimates. Figure 3 shows the point estimates and 95% predictive intervals by PRISM for the nowcasting during 2008–2016 in comparison to the true unemployment initial claims officially revealed a week later (in red). For 2008–2016, the actual coverage of the predictive interval is 96.6%, which is slightly higher than the nominal 95%. For longer forecasting horizons, the predictive intervals by PRISM also give coverage close to the nominal 95%: for the one-week-ahead, two-week-ahead and three-week-ahead forecasts, the actual coverage levels of the PRISM predictive intervals are, respectively, 93.9%, 95.4%, and 94.7% (detailed plots of the PRISM predictive intervals are provided in Supplementary Material A15).

3.2. Out-of-Sample Performance 2017–2019

To further assess PRISM's performance, we applied PRISM to produce out-of-sample forecasts of weekly unemployment
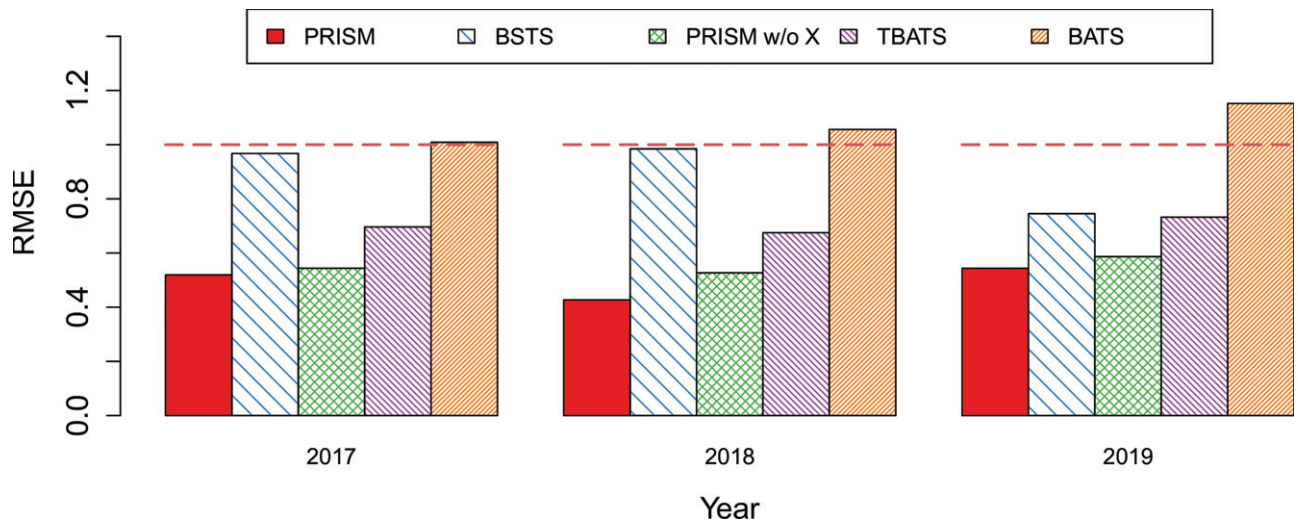
Figure 4. Yearly nowcasting performance of different methods from 2017 to 2019. RMSE is measured relative to the error of the naive method; a value above 1 indicates that the method performs worse than the naive method in that time period.

Table 3. Performance of different methods over 2017–2019 for four forecasting horizons: real-time, 1 week, 2 weeks, and 3 weeks ahead.

	Real-time	1 week	2 weeks	3 weeks
RMSE				
PRISM	**0.497**	**0.442**	**0.376**	**0.343**
PRISM w/o x_t	0.550	0.454	0.387	0.349
BSTS	0.921	–	–	–
BATS	1.064	0.982	0.907	0.832
TBATS	0.699	0.581	0.512	0.465
naive	1(27941)	1(34936)	1(41651)	1(46749)
MAE				
PRISM	**0.550**	**0.484**	**0.411**	0.368
PRISM w/o x_t	0.617	0.502	0.422	**0.365**
BSTS	0.941	-	-	-
BATS	1.155	0.972	0.907	0.816
TBATS	0.694	0.587	0.505	0.435
naive	1(19686)	1(24630)	1(29971)	1(35016)

NOTE: RMSE and MAE here are relative to the error of naive method; that is, the number reported is the ratio of the error of a given method to that of the naive method; the absolute RMSE and MAE of the naive method are reported in the parentheses. The boldface indicates the best performer for each forecasting horizon and each accuracy metric.

Table 4. p-values of the Diebold–Mariano test for prediction accuracy comparison between PRISM and the alternatives over 2017–2019 for four forecasting horizons: real-time, 1 week, 2 weeks, and 3 weeks ahead.

	Real-time	1 week	2 weeks	3 weeks
PRISM w/o x_t	2.74×10^{-9}	4.26×10^{-5}	1.25×10^{-5}	2.70×10^{-2}
BSTS	2.64×10^{-5}	–	–	–
BATS	3.80×10^{-7}	3.09×10^{-5}	2.40×10^{-5}	4.59×10^{-6}
TBATS	1.04×10^{-3}	2.79×10^{-3}	7.46×10^{-3}	2.48×10^{-3}

NOTE: The null hypothesis of the test is that PRISM and the alternative method in comparison have the same prediction accuracy in RMSE.

initial claims for the period of 2017–2019. Note that the PRISM methodology, including all the model specifications, was frozen at the end of 2016, so this evaluation is completely out of sample.

Table 3 summarizes the prediction errors of PRISM (both with and without Google data) in both RMSE and MAE in comparison with other benchmark methods. PRISM again shows consistent advantage over other benchmark methods in out-of-sample predictions. PRISM with Google data is leading across the board (except for the MAE of 3-week-ahead prediction where it virtually ties with PRISM without Google information). Notably, the relative errors compared with the naive method are quite stable over the years, similar to the results in 2007–2016. Breaking into each year (as shown in Figure 4), PRISM uniformly outperforms other methods in comparison and gives rather stable error reduction from the naive method over the years, both in- and out-of-sample. This figure together with Figure 2 shows that PRISM reduced around 50% error from the native method (in RMSE) year across year from 2007 to 2019. The statistical significance of PRISM's improved prediction

power is also verified by the Diebold–Mariano test in Table 4, where all the p-values are smaller than 3%. The consistent performance of PRISM both in the retrospective testing of 2007–2016 and the out-of-sample testing of 2017–2019 indicates the robustness and accuracy of PRISM over changes in economic environments and trends.

We also constructed point-wise predictive intervals based on the out-of-sample nowcasts in 2017–2019 (Figure 5). Compared with the actual unemployment data released one week later by the Department of Labor, the intervals by PRISM capture the actual numbers of unemployment initial claims in 97.1% of the weeks in 2017–2019, which is higher than the nominal 95% level. This again underscores the stability of the PRISM methodology.

3.3. Out-of-Sample Performance During COVID-19 Pandemic Period

The global shutdown due to the COVID-19 pandemic has heavily impacted the US economy and job market. In particular, the numbers of unemployment claims in the United States have skyrocketed to record-breaking levels with more than 40 millions people in total filing for initial claims since the start of the pandemic. This phenomenon has attracted significant attention from major news media and the general public (Casselman 2020; Cohen 2020a, 2020b). As the weekly unemployment claims remain "stubbornly high" (Casselman 2020), concerns
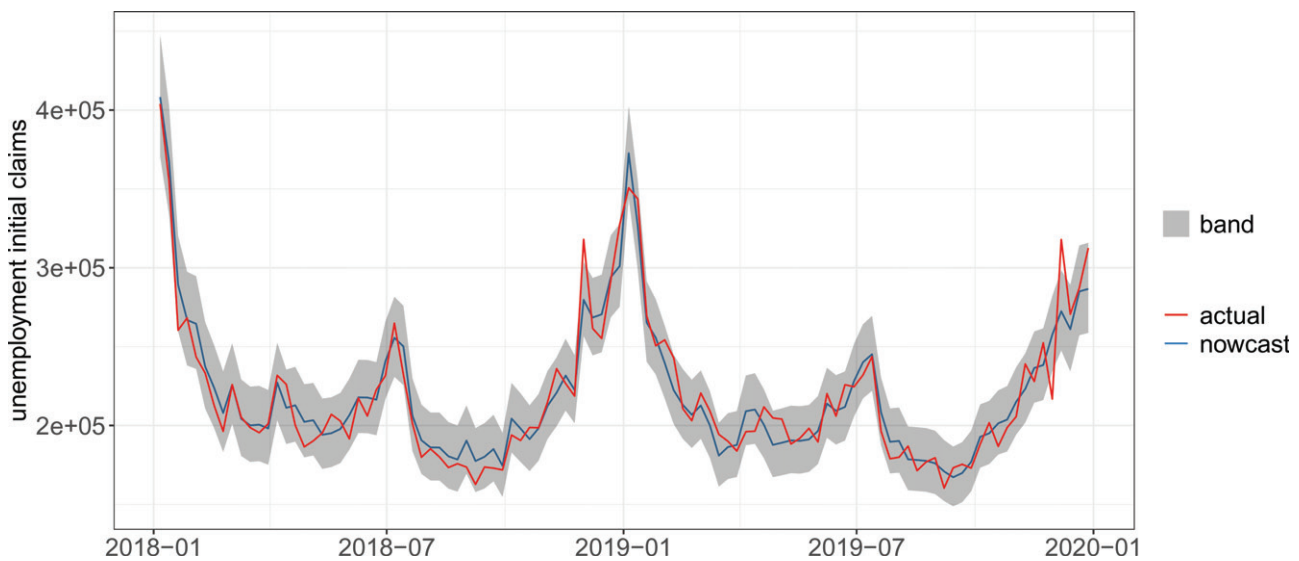
Figure 5. Predictive Interval of PRISM from 2017 to 2019. The shaded area corresponds to the 95% point-wise predictive interval of PRISM nowcasting. The blue curve is the point estimate of PRISM nowcasting. The red curve is the true unemployment initial claims. The actual coverage of the 95% PRISM predictive interval is 97.1% in 2017–2019.

Table 5. Performance of PRISM and benchmark methods during the COVID-19 pandemic period for real-time nowcasting.

	Mar 21–Apr 4	Apr 11–Apr 25	May 2–May 16	May 23–Jun 6	Jun 13–Jun 27	Jul 4–Jul 18
RMSE						
PRISM	0.684	**0.257**	0.607	**0.538**	**0.574**	**0.629**
BSTS	**0.543**	0.838	0.910	0.749	1.009	1.327
BATS	1.701	1.498	0.505	1.104	2.617	2.640
TBATS	1.862	0.432	**0.425**	1.497	4.496	1.817
naive	1 (2362454)	1 (932295)	1 (488192)	1 (232074)	1 (59761)	1 (105392)
MAE						
PRISM	0.784	**0.215**	0.462	**0.445**	**0.711**	**0.488**
BSTS	**0.539**	0.863	0.879	0.700	1.092	1.460
BATS	1.952	1.514	0.553	0.896	3.236	2.762
TBATS	2.136	0.363	**0.453**	1.651	5.940	1.818
naive	1 (1986663)	1 (898656)	1 (444600)	1 (206791)	1 (44883)	1 (95054)

NOTE: Evaluation period is broken down to 3-week windows. RMSE and MAE here are relative to the error of naive method; that is, the number reported is the ratio of the error of a given method to that of the naive method; the absolute RMSE and MAE of the naive method are reported in the parentheses. The boldface indicates the best performer for each forecasting horizon and each accuracy metric.

for significant layoffs and severe economic downturn persist. Accurate and reliable forecasting of near-future unemployment claims would thus provide very valuable insights into the trend of the general economy during such trying times. In light of this, we further applied PRISM to the out-of-sample data of the COVID-19 pandemic period to evaluate its performance and robustness to such unusual economic shock.

Table 5 and Figure 6 summarize the performance of PRISM's real-time nowcasting of the weekly unemployment claims in comparison with the other methods during the COVID pandemic period from March 21, 2020 to July 18, 2020. For close examination of the different methods, we break down the entire period into 3-week windows. During the first three weeks when the COVID-19 shutdown triggered the sudden and drastic rise in unemployment claims, both PRISM and BSTS picked up the signal rather quickly due to the input from the real-time Google Trends data that track the search of unemployment related query terms. Since April, PRISM began to show its advantage over the other methods as it adapted the predictive model to the "new regime," leading the chart in 5 out of 6 evaluation windows. It is worth pointing out that forecasting unemployment initial claims during this period is a very challenging

task as the unprecedented huge jump of unemployment claims drastically altered the pattern in the data (including the time-series pattern): we noted that PRISM is the only method that consistently outperforms the naive method throughout this COVID-19 period (the time-series based methods often performed worse than the naive method). This out-of-sample forecasting performance thus indicates the robustness of PRISM to unusual economic shocks and events, giving us more evidence of the model's reliability and accuracy. Further evaluation of PRISM and the benchmarks for longer-horizon predictions is presented in Supplementary Material A16, where PRISM also shows advantage in near-future predictions.

4. Discussion

The wide availability of data generated from the Internet offers great potential for predictive analysis and decision making. Our study on using Internet search data to forecast unemployment initial claims illustrates one such potential. The arrival of new data (sometimes in new forms) requires new methodology to analyze and use them. PRISM is an example where traditional
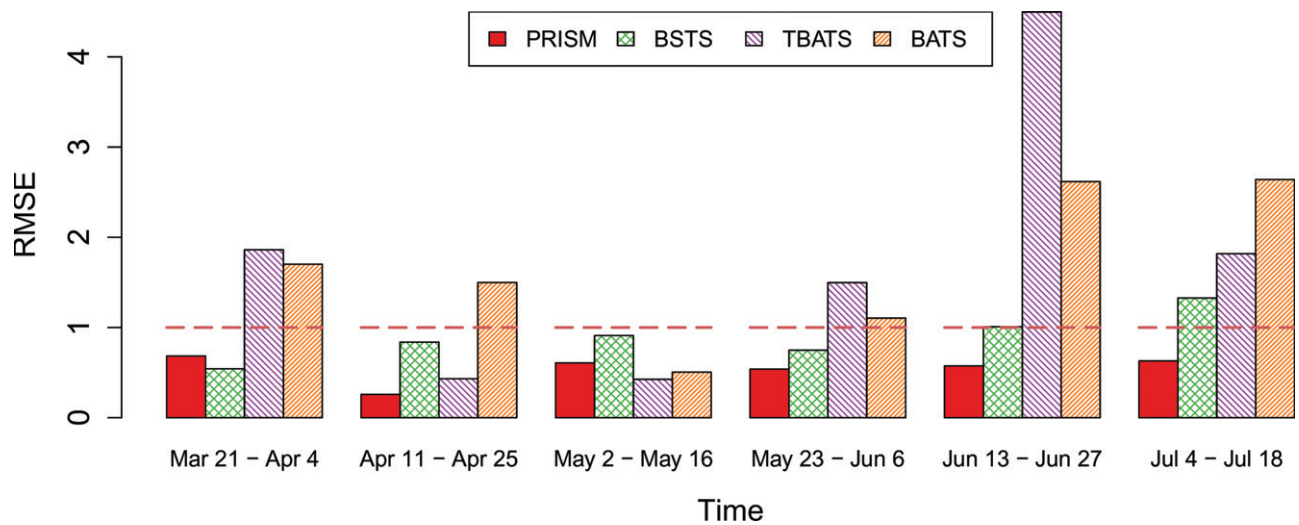
Figure 6. Nowcasting performance of different methods during the COVID-19 pandemic period (March 21, 2020 to July 18, 2020). RMSE is measured relative to the error of the naive method and evaluated in three-week windows; a value above 1 indicates that the method performs worse than the naive method in that time period.

statistical models are brought together with more recent statistical tools, such as L_1 regularization and dynamic training.

In this article we focus on using Internet search data to forecast unemployment initial claims weeks into the future. We introduced a novel method PRISM for forecasting time series with strong seasonality. PRISM is semiparametric and can be generally applied with or without exogenous variables. PRISM is motivated from a general state-space formulation that contains a variety of widely used time series models as special cases. The two stages of PRISM are easy to implement. The numerical evaluation shows that PRISM outperforms all alternatives in forecasting unemployment initial claim data for the time period of 2007–2019. We believe that the accurate and robust forecasts by PRISM would greatly benefit the public and the private sectors to assess and gauge the economic trends.

PRISM also demonstrates stable adaptability to unusual economic shocks such as the 2008–2009 financial crisis and the 2020 COVID-19 pandemic shutdown. The out-performance of PRISM relative to other methods are robust during long periods of economic expansion and during short periods of economic recession. In particular, during the 2008–2009 financial crisis and the COVID-19 pandemic period, we found that the real-time data from Google enables PRISM to quickly pick up the signal and the changes in data patterns and to provide insight on real-time and near-future economic trends. This gives us confidence that the unemployment forecasts given by PRISM would provide government agencies with much-needed information to react promptly and make well-informed decisions in the face of future economic and financial shocks.

The predictive power and advantage of PRISM mainly come from the following features: (1) dynamic model training based on a rolling window to account for changes in people's search pattern and changes in the relationship between Google search information and the targeted economic activity/index; (2) utilization of L_1 penalty to select the most relevant predictors and to filter out noisy and redundant information; (3) combination of nonparametric seasonality decomposition and penalized regression for greater flexibility and adaptability; (4) incorporation of real-time Google search information from multiple

related query terms to enhance prediction accuracy and robustness.

Although this article focuses on forecasting unemployment initial claims, PRISM can be generally used to forecast time series with complex seasonal patterns. The semiparametric approach of PRISM covers a wider range of time series models than traditional methods, as PRISM transforms the inference of a complicated class of state-space models into penalized regression of linear predictive models. Furthermore, dynamically fitting the predictive equations of PRISM addresses the time-varying relationship between the exogenous variables and the underlying time series. One interesting question for future study is to explore if we can extend PRISM to forecasting unemployment indicators in more specified industries such as construction, manufacturing, transportation, finance, and government or to forecasting other unemployment indicators such as nonfarm payrolls. Another direction for future study is to extend PRISM to predict unemployment indicators for different ethnic or demographic groups. Furthermore, it would also be of great future interests to see if PRISM can contribute to forecasting future breaks and macro-economic cycles.

We conclude this article with a few remarks on the detailed implementation of the PRISM method. We used real-time Internet search data from Google Trends, which provides publicly available data through subsampling and renormalization: a dataset undergoes subsampling (Google draws a small sample from its raw search data for a search query term) and renormalization (after the sampling, Google normalizes and rounds up the search volumes so that they become integers between 0 and 100 for each search query term) when downloaded. Due to the subsampling and renormalization, the search term volumes are noisy and variable (Yang, Santillana, and Kou 2015). The L_1 regularization adopted in PRISM has shown advantage in extracting the signals and reducing redundant information from Google search data (see Supplementary Materials A9 and A11). Furthermore, the dynamic training with rolling window accounts for changes in search engine algorithms, people's search patterns, economic trends and other patterns that change over time (Burkom, Murphy, and Shmueli 2007; Ning, Yang,

and Kou 2019). This is also evident in Figure A11, where each of the 25 candidate search terms has distinct patterns coming in and out of the dynamically fitted model throughout the time. The discount factor adopted also gives more weights on more recent data to capture more recent economic trends and Google search changes, which is similar to the data tapering idea proposed by Dahlhaus (1988, 1997) for improved performance in locally stationary time series. Our empirical analysis supports the effectiveness of the rolling window and discount factor (Supplementary Materials A6 and A7). Alternative frameworks for inferring time series models with state-space structures include the dynamic linear model (DLM) (Shumway and Stoffer 2017).

One limitation with PRISM arises from the data availability from Google Trends. The public available tool only provides up to 5-year range of weekly search data per download. Access to data in higher resolution and longer time span requires Google's permission to use its nonpublic API. Furthermore, in each downloaded batch, the search volume data are normalized by Google to the scale from 0 to 100 based on the queried search term and specific data range of that download. Thus, to avoid the variability due to the normalization and to ensure the consistency of results, we kept both the model training and the corresponding prediction within the same (downloaded) set of 5-year data. Furthermore, Google search data may not reflect the entire population of unemployed people, especially those who would not search online for employment information. Therefore, what we utilized is essentially the association between the search volume of related search terms and our target of unemployment claims for PRISM's prediction.

The R package `PRISM.forecast` that implements the PRISM method is available on CRAN at *https://CRAN.R-project.org/package=PRISM.forecast*. We also made the code available at *https://github.com/ryanddyi/prism*.

Funding

References

Ameen, J., and Harrison, P. (1984), "Discount Weighted Estimation," *Journal of Forecasting*, 3, 285–296. [1665]

Aoki, M. (1987), *State Space Modeling of Time Series*, Berlin, Heidelberg: Springer. [1663,1664]

Banbura, M., Giannone, D., Modugno, M., and Reichlin, L. (2013), "Now-Casting and the Real-Time Data Flow," in *Handbook of Economic Forecasting* (Vol. 2), eds. G. Elliott and A. Timmermann, Amsterdam: Elsevier, pp. 195–237. [1663]

Burkom, H. S., Murphy, S. P., and Shmueli, G. (2007), "Automated Time Series Forecasting for Biosurveillance," *Statistics in Medicine*, 26, 4202–4218. [1671]

Casselman, B. (2020), "Rise in Unemployment Claims Signals an Economic Reversal," *The New York Times*. [1669]

Chen, H., Chiang, R. H., and Storey, V. C. (2012), "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly*, 36, 1165–1188. [1662]

Choi, H., and Varian, H. (2012), "Predicting the Present With Google Trends," *Economic Record*, 88, 2–9. [1662]

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990), "STL: A Seasonal-Trend Decomposition," *Journal of Official Statistics*, 6, 3–73. [1665]

Cohen, P. (2020a), "Rise in Unemployment Claims Signals an Economic Reversal," *The New York Times*. [1669]

——— (2020b), "'Still Catching Up': Jobless Numbers May Not Tell Full Story," *The New York Times*. [1669]

Dahlhaus, R. (1988), "Small Sample Effects in Time Series Analysis: A New Asymptotic Theory and a New Estimate," *The Annals of Statistics*, 16, 808–841. [1672]

——— (1997), "Fitting Time Series Models to Nonstationary Processes," *The Annals of Statistics*, 25, 1–37. [1672]

D'Amuri, F., and Marcucci, J. (2017), "The Predictive Power of Google Searches in Forecasting US Unemployment," *International Journal of Forecasting*, 33, 801–816. [1666,1667]

De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011), "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing," *Journal of the American Statistical Association*, 106, 1513–1527. [1666]

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263. [1667]

Durbin, J., and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods*, Oxford: Oxford University Press. [1663]

Einav, L., and Levin, J. (2014), "The Data Revolution and Economic Analysis," *Innovation Policy and the Economy*, 14, 1–24. [1662]

Ettredge, M., Gerdes, J., and Karuga, G. (2005), "Using Web-Based Search Data to Predict Macroeconomic Statistics," *Communications of the ACM*, 48, 87–92. [1662]

Giannone, D., Reichlin, L., and Small, D. (2008), "Nowcasting: The Real-Time Informational Content of Macroeconomic Data," *Journal of Monetary Economics*, 55, 665–676. [1662]

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009), "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, 457, 1012–1014. [1662]

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. (2010), "Predicting Consumer Behavior With Web Search," *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17486–17490. [1662]

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press. [1663,1664]

Harvey, A., and Koopman, S. J. (1993), "Forecasting Hourly Electricity Demand Using Time-Varying Splines," *Journal of the American Statistical Association*, 88, 1228–1236. [1663]

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008), *Forecasting With Exponential Smoothing: The State Space Approach*, Berlin, Heidelberg: Springer. [1663,1664,1665]

Khoury, M. J., and Ioannidis, J. P. (2014), "Big Data Meets Public Health," *Science*, 346, 1054–1055. [1662]

Kim, G.-H., Trimi, S., and Chung, J.-H. (2014), "Big-Data Applications in the Government Sector," *Communications of the ACM*, 57, 78–85. [1662]

Li, X. (2016), "Nowcasting With Big Data: Is Google Useful in Presence of Other Information," Policy Research Working Paper. [1662]

Lindoff, B. (1997), "On the Optimal Choice of the Forgetting Factor in the Recursive Least Squares Estimator," Technical Report, Lund Institute of Technology. [1666]

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, Washington, DC: McKinsey & Company. [1662]

McAfee, A., and Brynjolfsson, E. (2012), "Big Data: The Management Revolution," *Harvard Business Review*, 90, 60–68. [1662]

Murdoch, T. B., and Detsky, A. S. (2013), "The Inevitable Application of Big Data to Health Care," *Journal of the American Medical Association*, 309, 1351–1352. [1662]

Ning, S., Yang, S., and Kou, S. (2019), "Accurate Regional Influenza Epidemics Tracking Using Internet Search Data," *Scientific Reports*, 9, 5238. [1662,1672]

Ord, J. K., Koehler, A. B., and Snyder, R. D. (1997), "Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models," *Journal of the American Statistical Association*, 92, 1621–1629. [1664]

Preis, T., Moat, H. S., and Stanley, H. E. (2013), "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Scientific Reports*, 3, 1684. [1662]

Risteski, D., and Davcev, D. (2014), "Can We Use Daily Internet Search Query Data to Improve Predicting Power of EGARCH Models for Financial Time Series Volatility," in *Proceedings of the International Conference on Computer Science and Information Systems (ICSIS'2014)*, October 17–18, 2014, Dubai, United Arab Emirates. [1662]

Scott, S. L., and Varian, H. R. (2013), "Bayesian Variable Selection for Nowcasting Economic Time Series," Technical Report, National Bureau of Economic Research. [1663]

——— (2014), "Predicting the Present With Bayesian Structural Time Series," *International Journal of Mathematical Modelling and Numerical Optimisation*, 5, 4–23. [1663,1666]

Shumway, R. H., and Stoffer, D. S. (2017), *Time Series Analysis and Its Applications: With R Examples*, Cham: Springer. [1672]

Siegel, E. (2013), *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*, Hoboken, NJ: Wiley. [1662]

Stuart, A., and Kendall, M. G. (1963), *The Advanced Theory of Statistics*, London: Griffin. [1665]

Taylor, J. W. (2010), "Exponentially Weighted Methods for Forecasting Intraday Time Series With Multiple Seasonal Cycles," *International Journal of Forecasting*, 26, 627–646. [1665]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [1665]

Welch, I., and Goyal, A. (2008), "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *The Review of Financial Studies*, 21, 1455–1508. [1667]

Wu, L., and Brynjolfsson, E. (2015), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Sales," in *Economic Analysis of the Digital Economy*, University of Chicago Press, pp. 89–118. [1662]

Yang, S., Kou, S. C., Lu, F., Brownstein, J. S., Brooke, N., and Santillana, M. (2017), "Advances in Using Internet Searches to Track Dengue," *PLoS Computational Biology*, 13, e1005607. [1662]

Yang, S., Ning, S., and Kou, S. (2021), "Use Internet Search Data to Accurately Track State-Level Influenza Epidemics," *Scientific Reports*, 11, 4023. [1662]

Yang, S., Santillana, M., and Kou, S. C. (2015), "Accurate Estimation of Influenza Epidemics Using Google Search Data via ARGO," *Proceedings of the National Academy of Sciences of the United States of America*, 112, 14473–14478. [1662,1671]

Supplementary Material

Details of the methodology and derivation of PRISM are presented as follows. First, the exact search query terms used in our study, which were identified from Google Trends, are presented. Second, the general state-space formulation that motivates PRISM is presented together with a few widely used special cases. Third, the predictive distribution for forecasting is described in detail. Fourth, the robustness of PRISM to the choice of the seasonal decomposition method and to the choice of the discount factor is presented. Fifth, the detailed proof of the mathematical result is provided.

A1. Internet Search Data from Google

The real-time Internet search data we used were from Google Trends (www.google.com/trends). The search query terms that we used in our study were also identified from the Google Trends tool. One feature of Google Trends is that, in addition to the time series of a specific term (or a general topic), it also returns the top query terms that are most highly correlated with the specific term. In our study, we used a list of top 25 Google search terms that are the most highly correlated with the term "unemployment". Table A1 lists these 25 terms, which were generated by Google Trends on January 11, 2018. Figure A1, the upper panel, illustrates the Google Trend time series of several search query terms in a 5-year span. Comparing these time series to the lower panel of Figure A1, which shows the unemployment initial claims in the same time period, it is evident that the former provides noisy signal about the latter. On the Google Trends site, the weekly data are available for at most a 5-year span in a query, and it would be automatically transformed to monthly data if one asks for more than 5 years. To model and forecast the weekly unemployment claims for the entire period of 2007-2016, we downloaded separate weekly data sets from Google Trends, covering 2004-2008, 2006-2010, 2008-2012, 2010-2014 and 2012-2016, respectively.

To avoid the variability due to the normalization and to ensure the coherence of the model, for each week, we kept both the training data and the data used for prediction within the same 5-year span of data downloaded. For each search term, we downloaded Google Trends data based on the same 5-year range and the multivariate x_t. Then we trained the model and made predictions based on a rolling window of 3 years. So for the same set of data downloaded with the range of 2004-2008, we are able to make predictions for weeks in 2007-2008; similarly, the data covering 2006-2010 will give predictions for weeks in 2009-2010, etc. See Figure A2 for an illustration.

Table A1: Top 25 nationwide search query terms associated with the term "unemployment" generated by Google Trends as of January 11, 2018.

|		
unemployment	unemployment benefits	unemployment rate
unemployment office	pa unemployment	claim unemployment
ny unemployment	nys unemployment	ohio unemployment
unemployment florida	unemployment extension	texas unemployment
nj unemployment	unemployment number	file unemployment
unemployment insurance	california unemployment	unemployed
unemployment oregon	new york unemployment	indiana unemployment
unemployment washington	unemployment wisconsin	unemployment online
unemployment login		
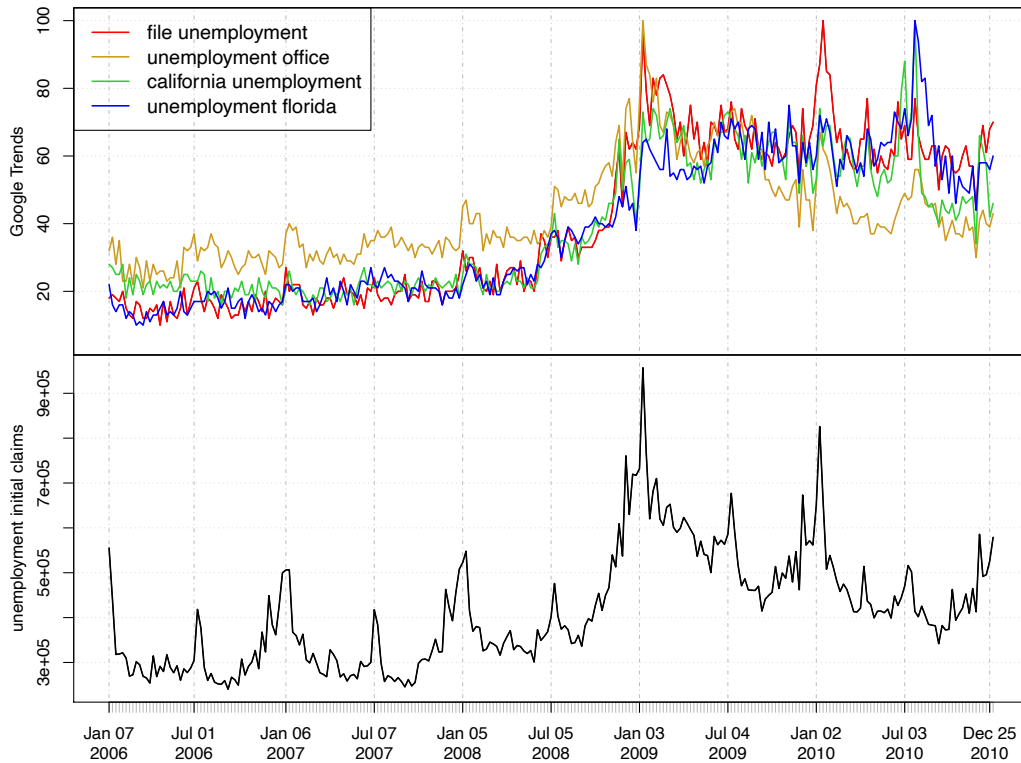
Figure A1: The upper panel shows the Google Trends data of four unemployment related search queries in 2006-2010. The lower panel shows the weekly unemployment initial claims data in the same time period.
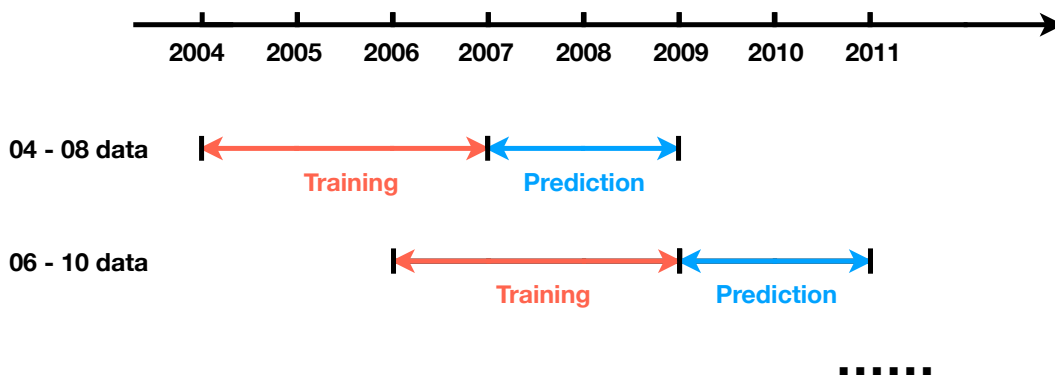

Figure A2: Illustration of the rolling-window prediction scheme based on 5-year span of Google Trends data.

A2. Our State-Space Formulation and Special Cases

Our state-space model, Eq. (2.2) in the main text, contains a variety of widely used time series models, including structural time series models and additive innovation state-space models. Under this general formulation, a specific parametric model can be obtained by specifying the state-space models for z_t and γ_t along with their dependence structure $\boldsymbol{H}$.

A2.1. Special case 1: seasonal AR model

The following AR model with seasonal pattern is a special case: modeling z_t as an autoregressive process with lag N and assuming a dummy variable formulation with period S for the seasonal component γ_t:

$$
\begin{aligned}
y_t &= z_t + \gamma_t, \\
z_t &= \mu_z + \sum_{j=1}^{N} \alpha_j z_{t-j} + \eta_t, \quad \eta_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\eta^2) \\
\gamma_t &= - \sum_{j=1}^{S-1} \gamma_{t-j} + \omega_t, \quad \omega_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\omega^2)
\end{aligned}
\tag{A1}
$$

The dummy variable model for the seasonal component implies that sum of the seasonal components over the S periods, $\sum_{j=0}^{S-1} \gamma_{t-j}$, has mean zero and variance σ_ω^2. In the seasonal AR model (A1), each time series block of $\{z_{(t-N+1):t}\}_{t \geq N}$ and $\{\gamma_{(t-S+2):t}\}_{t \geq (S-1)}$ evolves as a Markov Chain. Under our general state-space model, Eq. (2.2) in the main text, if we set $\boldsymbol{h}_t = (1, z_t, z_{t-1}, \ldots, z_{t-N+1})$ and $\boldsymbol{s}_t = (\gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-S+2})$, then it reduces to the seasonal AR model (A1).

A2.2. Special case 2: structural time series models

The basic structural model assumes that a univariate time series is the sum of trend, seasonal and irregular components, each of which follows an independent stochastic process (Harvey 1989). The model is

$$
y_t = \mu_t + \gamma_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2),
\tag{A2}
$$

where μ_t is the trend component, and γ_t and ϵ_t are the seasonal and irregular components, respectively. The trend is often specified by a local level model

$$
\begin{aligned}
\mu_t &= \mu_{t-1} + \delta_t + \eta_t, & \eta_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\eta^2), & \tag{A3a} \\
\delta_t &= \delta_{t-1} + \zeta_t, & \zeta_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\zeta^2), & \tag{A3b}
\end{aligned}
$$

where μ_t is the level and δ_t is the slope. η_t and ζ_t are assumed mutually independent. For time series with S periods, the seasonal component can be specified through the seasonal dummy variable model

$$
\gamma_t = - \sum_{j=1}^{S-1} \gamma_{t-j} + \omega_t, \qquad \omega_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_\omega^2).
\tag{A4}
$$

which is the same as the seasonal component in the seasonal AR model (A1). Alternatively, the seasonal pattern can be modeled by a set of trigonometric terms at seasonal frequencies $\lambda_j = 2\pi j/S$ (Harvey 1989):

$$\gamma_t = \sum_{j=1}^{[S/2]} \gamma_{j,t}, \tag{A5a}$$

$$\begin{pmatrix} \gamma_{j,t} \\ \gamma_{j,t}^* \end{pmatrix} = \begin{pmatrix} \cos\lambda_j & \sin\lambda_j \\ -\sin\lambda_j & \cos\lambda_j \end{pmatrix} \begin{pmatrix} \gamma_{j,t-1} \\ \gamma_{j,t-1}^* \end{pmatrix} + \begin{pmatrix} \omega_{j,t} \\ \omega_{j,t}^* \end{pmatrix}, \tag{A5b}$$

where $\omega_{j,t}$ and $\omega_{j,t}^*$, $j = 1,\ldots,[S/2]$, are independent and normally distributed with common variance σ_ω^2.

Under our general state-space model, Eq. (2.2) of the main text, if we take $z_t = \mu_t + \epsilon_t$ and $h_t = (\mu_t, \delta_t)$, then it specializes to structural time series models. In particular, for the dummy variable seasonality of (A4), s_t in the general model corresponds to $s_t = (\gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-S+2})$; and for the trigonometric seasonality of (A5), s_t in the general model corresponds to $s_t = (\gamma_{1,t}, \ldots, \gamma_{[S/2],t}, \gamma_{1,t}^*, \ldots, \gamma_{[S/2],t}^*)$.

A2.3. Special case 3: additive innovations state-space models

An alternative to structural time series models, which have multiple sources of error, innovation state-space model (Aoki 1987), where the same error term appears in each equation, is also popular. These innovation state-space models underlie exponential smoothing methods, which are widely used in time series forecasting and have been proven optimal under many specifications of the innovation state-space model (Ord et al. 1997; Hyndman et al. 2008). Among exponential smoothing methods, Holt-Winters' method (Holt 1957; Winters 1960) is developed to capture both trend and seasonality, and it postulates a model specification similar to the basic structural model (A2)- (A4). In particular, Holt-Winters' additive method is

$$\begin{aligned} y_t &= \mu_{t-1} + \delta_{t-1} + \gamma_{t-S} + \epsilon_t, & \text{(A6a)} \\ \mu_t &= \mu_{t-1} + \delta_{t-1} + \alpha\epsilon_t, & \text{(A6b)} \\ \delta_t &= \delta_{t-1} + \beta\epsilon_t, & \text{(A6c)} \\ \gamma_t &= \gamma_{t-S} + \omega\epsilon_t, & \text{(A6d)} \end{aligned}$$

where the components μ_t, δ_t and γ_t represent level, slope and seasonal components of time series, and $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is the only source of error. Since Eq. (A6a) can be rewritten as

$$y_t = \mu_t + \gamma_t + (1 - \alpha - \omega)\epsilon_t,$$

we observe that model (A6) is special case of our general formulation with $z_t = \mu_t + (1-\alpha-\omega)\epsilon_t$, $h_t = (\mu_t, \delta_t)$ and $s_t = (\gamma_t, \gamma_{t-1}, \ldots, \gamma_{t-S+1})$ in Eq. (2.2) of the main text. The Holt-Winters model is among a collection of innovation state-space models that are summarized in Hyndman et al. (2008) using the triplet (E, T, S), which represents model specification for the three components: error, trend and seasonality. For instance, (A6) is also referred to as local additive seasonal model or ETS(A,A,A), where A stands for additive. Our general state-space formulation, Eq. (2.2) in the main text, also incorporates many useful model extensions as special cases, including the

4

damped trend (Gardner Jr and McKenzie 1985) and multiple seasonal patterns (Gould et al. 2008; De Livera et al. 2011). For example, the damped trend double seasonal model extends model (A6) to include a factor $\phi \in [0, 1)$ and a second seasonal component as follows:

$$
\begin{aligned}
y_t &= \mu_{t-1} + \phi \delta_{t-1} + \gamma^{(1)}_{t-S_1} + \gamma^{(2)}_{t-S_2} + \epsilon_t, \\
\mu_t &= \mu_{t-1} + \phi \delta_{t-1} + \alpha \epsilon_t, \\
\delta_t &= \phi \delta_{t-1} + \beta \epsilon_t, \\
\gamma^{(1)}_t &= \gamma^{(1)}_{t-S_1} + \omega_1 \epsilon_t, \\
\gamma^{(2)}_t &= \gamma^{(1)}_{t-S_2} + \omega_2 \epsilon_t.
\end{aligned}
\tag{A7}
$$

Our general model contains this extended model, where $z_t = \mu_t + (1 - \alpha - \omega_1 - \omega_2)\epsilon_t$, $\gamma_t = \gamma^{(1)}_t + \gamma^{(2)}_t$, $\boldsymbol{h}_t = (\mu_t, \delta_t)$ and $\boldsymbol{s}_t = (\gamma^{(1)}_t, \dots, \gamma^{(1)}_{t-S_1+1}, \gamma^{(2)}_t, \dots, \gamma^{(2)}_{t-S_2+1})$.

A2.4. Motivation of the general formulation

The motivation of our general state-space formulation is to collectively consider all possible models under it and to *semi-parametrically* obtain the prediction under this large class of models. In comparison, traditional time series studies often rely on parameter estimation of specified models such as those highlighted in the previous subsections. For instance, exponential smoothing is tailored for computing the likelihood and obtaining maximum likelihood estimates of the innovation state-space models. For other parametric models with multiple sources of error, their likelihood might be evaluated by the Kalman filter, but the parameter estimation can be difficult in many cases. In the traditional parametric times series model setting, model selections are often applied by optimizing certain selection criteria (e.g. AIC or BIC), but when the class of models under consideration become really large such as Eq. (2.2) of the main text, traditional model selection methods encounter serious challenges (as they lack scalability) to operate on such a wide range of models. As a consequence, traditional parametric time series models often consider a much smaller collection of models compared to Eq. (2.2) of the main text. The cost of focusing on a small class of models is that the forecasting accuracy can substantially suffer as the risk of model misspecification is high.

To relieve these challenges and improve the performance of forecasting, we use our general state-space formulation to motivate a semi-parametric method for forecasting time series. We derive and study a linear predictive model that is coherent with all possible models under Eq. (2.2) of the main text. With forecasting as our main goal, we essentially transform the question from the inference of a complicated class of state-space models into penalized regression and forecasting based on a linear prediction formulation.

A3. Predictive Distributions for Forecasting

Under our general state-space model – Eq. (2.2) and Eq. (2.3) of the main text – given the historical data $\{y_{1:(t-1)}\}$ and contemporaneous exogenous time series $\{\boldsymbol{x}_{t_0:t}\}$, the predictive distribution for forecasting y_{t+l} ($l \geq 0$) at time t would be $p(y_{t+l} \mid y_{1:(t-1)}, \boldsymbol{x}_{t_0:t})$. In PRISM, we consider the

predictive distribution of y_t by further conditioning on the latent seasonal component $\{\gamma_t\}$:

$$p(y_{t+l} \mid y_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t}). \tag{A1}$$

Note that since $z_t = y_t - \gamma_t$ for all t, $z_{1:(t-1)}$ is known given $y_{1:(t-1)}$ and $\gamma_{1:(t-1)}$. The advantage of working on (A1) is that we can establish a universal representation of the predictive distribution as given by the next proposition.

Proposition 1. *Under our model — Eq. (2.2) and (2.3) of the main text — y_{t+l} ($l \geq 0$) conditioning on $\{z_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t}\}$ follows a normal distribution with the conditional mean $E(y_{t+l} \mid z_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t})$ linear in $z_{1:(t-1)}$, $\gamma_{1:(t-1)}$ and $\boldsymbol{x}_t$.*

As a partial result that leads to Proposition 1, we have, without the exogenous variables, the conditional distribution of $y_{t+l} \mid z_{1:(t-1)}, \gamma_{1:(t-1)}$ is normal with mean linear in $z_{1:(t-1)}$ and $\gamma_{1:(t-1)}$.

Based on Proposition 1, we can represent $p\left(y_{t+l} \mid z_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t}\right)$ as

$$y_{t+l} = \mu_t^{(l)} + \sum_{j=1}^{t-1} \alpha_{j,t}^{(l)} z_{t-j} + \sum_{j=1}^{t-1} \delta_{j,t}^{(l)} \gamma_{t-j} + \sum_{i=1}^{p} \beta_{i,t}^{(l)} x_{i,t} + \epsilon_t, \quad \epsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_{t,l}^2), \tag{A2}$$

where $\mu_t^{(l)}$, $\alpha_{j,t}^{(l)}$, $\delta_{j,t}^{(l)}$, $\beta_{i,t}^{(l)}$ and $\sigma_{t,l}^2$ are fixed but unknown constants that are determined by original parameters $\boldsymbol{\theta}$ and the initial values of the state vectors.

A4.   Robustness to the Choice of Seasonal Decomposition Method

We compare the performance of PRISM with two seasonal decomposition methods: STL and the classic additive decomposition. Both methods decompose target time series y_t into the trend component T_t, the seasonal component S_t and the irregular component R_t:

$$y_t = T_t + S_t + R_t.$$

In the classic additive decomposition, the estimated trend component $\hat{T}_t$ is calculated from the moving average of $\{y_t\}$. The seasonal component S_t for a certain week is assumed to be the same in each period, and is estimated by the average of the detrended value, $y_t - \hat{T}_t$ for the specific week. For example, assuming 52 weeks in a year, the seasonal index for the fifth week is the average of all the detrended fifth week values in the data.

In contrast, STL relies on a sequence of applications of loess smoother to generate the seasonal and trend components. The implementation of STL involves two loops. In the outer loop, robustness weights are assigned to each data point to reduce the effects of outliers, while the inner loop iteratively updates the trend and seasonal components. In the inner loop iteration, the seasonal components are updated using detrended time series similar to classic additive decomposition, and the trend components are calculated by loess smoothing of the deseasonalized time series.

Both STL and the classic additive seasonal decomposition are options in the R package of PRISM with STL being the default setting. Table A2 describes the performance of PRISM with the two different methods. The numerical result of PRISM is quite robust to the choice of the seasonal decomposition method with STL providing slightly better overall result.

Table A2: The performance of PRISM with two different seasonal decomposition methods: STL and additive decomposition. RMSE and MAE are measured relative to the respective error of the naive method. The boldface highlights the better performance for each metric and forecasting horizon.

	real-time	forecast 1 wk	forecast 2 wk	forecast 3 wk
RMSE				
additive decomposition	**0.496**	0.497	0.462	**0.460**
STL decomposition	0.498	**0.492**	**0.453**	0.467
MAE				
additive decomposition	0.543	0.538	0.482	**0.458**
STL decomposition	**0.542**	**0.534**	**0.479**	0.465

A5.   Effect of the Discount Factor

We tested the effect of the discount factor w. Lindoff (1997) suggested setting w between 0.95 and 0.995 for practical applications. Table A3 shows the performance of PRISM with different $w \in [0.95, 1]$ (note that $w = 1$ corresponds to no discount). The performance of PRISM is seen to be quite robust for the different choices of w between $[0.95, 0.995]$. The discount factor w is an option in the R package of PRISM, and the default value is set to be 0.99. We can also see that comparing with no discounting ($w = 1$), our default choice of discount factor can provide up to 3% more error reduction in terms of the relative errors to the naive method. This motivates us to incorporate the discount factor in our model.

A6.   Proof of Proposition 1

We first prove the case for $l = 0$. Let $r_t = (z_t, \gamma_t)'$. Since $y_t = z_t + \gamma_t$ for all t, $y_t \mid y_{1:(t-1)}, \gamma_{1:(t-1)}$ is equivalent to $z_t + \gamma_t \mid z_{1:(t-1)}, \gamma_{1:(t-1)}$. By treating r_t as a 2-dimensional observable and $\alpha_t = (h_t', s_t')'$ as the state vector, we can rewrite Eq. (2.2) of the main text as

$$r_t = \Phi' \alpha_t + \xi_t$$
$$\alpha_t = \Lambda \alpha_{t-1} + \tau_t$$

(A1)

where $\Phi = \begin{bmatrix} w & 0 \\ 0 & v \end{bmatrix}$, $\Lambda = \begin{bmatrix} F & O \\ O & P \end{bmatrix}$ and $(\xi_t', \tau_t')' \overset{iid}{\sim} \mathcal{N}(0, H)$.

Denote $r_{1:t} = (r_1', \ldots, r_t')'$ and $\alpha_{1:t} = (\alpha_1', \ldots, \alpha_t')'$. According to the property of Gaussian linear state-space model, $r_{1:t}$ and $\alpha_{1:t}$ jointly follows a multivariate normal distribution. Therefore, the sub-vector $r_{1:t}$ is also normal, and $r_t \mid r_{1:(t-1)}$ follows a bivariate normal distribution with mean linear in $r_{1:(t-1)}$, i.e.

$$r_t \mid r_{1:(t-1)} \sim \mathcal{N}(\Gamma_t r_{1:(t-1)}, \Sigma_t),$$

(A2)

where Γ_t and Σ_t are determined by Φ, Λ and H. For any given parameters, the above distribution can be numerically evaluated through Kalman filter. Here, we focus only on the general analytical

Table A3: The performance of PRISM for the discount factor $w \in [0.95, 1]$. RMSE and MAE are measured relative to the respective error of the naive method. The boldface highlights the best performance for each metric and forecasting horizon.

	real-time	forecast 1 wk	forecast 2 wk	forecast 3 wk
RMSE				
$w = 1$	0.515	0.513	0.470	0.465
$w = 0.995$	0.512	0.501	0.463	0.462
$w = 0.99$	0.495	0.491	**0.457**	**0.461**
$w = 0.985$	**0.494**	0.486	0.462	0.466
$w = 0.98$	0.506	0.489	0.471	0.472
$w = 0.975$	0.500	0.485	0.471	0.468
$w = 0.97$	0.513	0.488	0.475	0.495
$w = 0.965$	0.516	0.488	0.485	0.479
$w = 0.96$	0.519	0.492	0.486	0.483
$w = 0.955$	0.525	0.502	0.502	0.497
$w = 0.95$	0.540	**0.484**	0.482	0.491
MAE				
$w = 1$	0.571	0.568	0.507	0.476
$w = 0.995$	0.556	0.546	0.487	0.467
$w = 0.99$	0.545	0.532	**0.478**	0.461
$w = 0.985$	**0.538**	0.522	0.480	**0.460**
$w = 0.98$	0.543	0.517	0.484	**0.460**
$w = 0.975$	0.540	**0.516**	0.481	0.461
$w = 0.97$	0.549	0.519	0.488	0.482
$w = 0.965$	0.555	0.523	0.495	0.465
$w = 0.96$	0.552	0.524	0.496	0.468
$w = 0.955$	0.563	0.533	0.511	0.477
$w = 0.95$	0.572	0.521	0.496	0.474

formulation. Following (A2) and $y_t = \mathbf{1}' \mathbf{r}_t$, we have

$$y_t \mid \mathbf{r}_{1:(t-1)} \sim \mathcal{N}(\mathbf{1}' \mathbf{\Gamma}_t \mathbf{r}_{1:(t-1)}, \mathbf{1}' \mathbf{\Sigma}_t \mathbf{1}).$$

Thus, given $\mathbf{r}_{1:(t-1)}$, or equivalently $z_{1:(t-1)}$ and $\gamma_{1:(t-1)}$, y_t has a univariate normal distribution with mean linear in $z_{1:(t-1)}$ and $\gamma_{1:(t-1)}$.

When taking the exogenous variable $\mathbf{x}_t$ into account, we have $p(y_t \mid \mathbf{r}_{1:(t-1)}, \mathbf{x}_{t_0:t}) \propto p(\mathbf{x}_t \mid y_t) p(y_t \mid \mathbf{r}_{1:(t-1)})$. Since

$$p(\mathbf{x}_t \mid y_t) \propto \exp\left(-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}_x - y_t \boldsymbol{\beta})' \mathbf{Q}^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_x - y_t \boldsymbol{\beta}) \right), \tag{A3}$$

it follows that

$$p(y_t \mid \mathbf{r}_{1:(t-1)}, \mathbf{x}_{t_0:t}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}_x - y_t \boldsymbol{\beta})' \mathbf{Q}^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_x - y_t \boldsymbol{\beta}) - \frac{1}{2}(\mathbf{1}' \mathbf{\Sigma}_t \mathbf{1})^{-1}(y_t - \mathbf{1}' \mathbf{\Gamma}_t \mathbf{r}_{1:(t-1)})^2 \right).$$

Hence, $y_t \mid \mathbf{r}_{1:(t-1)}, \mathbf{x}_{t_0:t}$ is normal, since the above equation is an exponential function of a

quadratic form of y_t. By reorganizing the terms, we have

$$E\left(y_t \mid \boldsymbol{r}_{1:(t-1)}, \boldsymbol{x}_{t_0:t}\right) = \left(\left(\mathbf{1}'\boldsymbol{\Sigma}_t\mathbf{1}\right)^{-1} + \boldsymbol{\beta}'\boldsymbol{Q}^{-1}\boldsymbol{\beta}\right)^{-1} \left(\left(\mathbf{1}'\boldsymbol{\Sigma}_t\mathbf{1}\right)^{-1}\mathbf{1}'\boldsymbol{\Gamma}_t\boldsymbol{r}_{1:(t-1)} + \boldsymbol{\beta}'\boldsymbol{Q}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu}_x)\right)$$

and

$$Var\left(y_t \mid \boldsymbol{r}_{1:(t-1)}, \boldsymbol{x}_{t_0:t}\right) = \left(\left(\mathbf{1}'\boldsymbol{\Sigma}_t\mathbf{1}\right)^{-1} + \boldsymbol{\beta}'\boldsymbol{Q}^{-1}\boldsymbol{\beta}\right)^{-1}.$$

Therefore, $y_t \mid z_{1:(t-1)}, \gamma_{1:(t-1)}, \boldsymbol{x}_{t_0:t}$ has a normal distribution with mean linear in $z_{1:(t-1)}, \gamma_{1:(t-1)}$ and $\boldsymbol{x}_t$.

Next, we prove the case for $l \geq 1$. We consider the following predictive distribution

$$p\left(y_{t+l} \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \quad \propto \quad \int p\left(y_{t+l}, \boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \mathrm{d}\boldsymbol{r}_t$$

$$\propto \quad \int p\left(y_{t+l} \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:t}\right) p\left(\boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \mathrm{d}\boldsymbol{r}_t.$$

Since $\boldsymbol{x}_{t_0:t}$ is independent of y_{t+l} conditional on $y_{1:t}$, $p(y_{t+l} \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:t}) = p(y_{t+l} \mid \boldsymbol{r}_{1:t})$. Similarly, $\boldsymbol{x}_{t_0:(t-1)}$ is independent of $\boldsymbol{r}_t$ conditional on $\boldsymbol{r}_{1:(t-1)}$, which implies $p(\boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}) = p(\boldsymbol{r}_t \mid \boldsymbol{x}_t, \boldsymbol{r}_{1:(t-1)})$. Note that $p\left(\boldsymbol{r}_t \mid \boldsymbol{x}_t, \boldsymbol{r}_{1:(t-1)}\right) \propto p(\boldsymbol{x}_t \mid \boldsymbol{r}_t)p(\boldsymbol{r}_t \mid \boldsymbol{r}_{1:(t-1)})$. Thus, we have

$$p\left(y_{t+l}, \boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \quad \propto \quad p\left(y_{t+l} \mid \boldsymbol{r}_{1:t}\right) p\left(\boldsymbol{r}_t \mid \boldsymbol{x}_t, \boldsymbol{r}_{1:(t-1)}\right)$$

$$\propto \quad p\left(y_{t+l} \mid \boldsymbol{r}_{1:t}\right) p(\boldsymbol{x}_t \mid \boldsymbol{r}_t)p(\boldsymbol{r}_t \mid \boldsymbol{r}_{1:(t-1)}).$$

In the first part of the proof, we have learned that $\boldsymbol{r}_{1:(t+l)}$ is multivariate normal. Similar to (A2), we can write $\boldsymbol{r}_{t+l} \mid \boldsymbol{r}_{1:t}$ as

$$\boldsymbol{r}_{t+l} \mid \boldsymbol{r}_{1:t} \sim \mathcal{N}(\boldsymbol{\Gamma}_{t,l}\boldsymbol{r}_{1:t}, \boldsymbol{\Sigma}_{t,l}), \tag{A4}$$

where $\boldsymbol{\Gamma}_{t,l}$ and $\boldsymbol{\Sigma}_{t,l}$ are determined by $\boldsymbol{\Phi}, \boldsymbol{\Lambda}$ and $\boldsymbol{H}$. Hence, $y_{t+l} \mid \boldsymbol{r}_{1:t} \sim \mathcal{N}(\mathbf{1}'\boldsymbol{\Gamma}_{t,l}\boldsymbol{r}_{1:t}, \mathbf{1}'\boldsymbol{\Sigma}_{t,l}\mathbf{1})$. Combining the above results with (A3), we have

$$p\left(y_{t+l}, \boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \propto \exp\left(-\frac{1}{2}(\mathbf{1}'\boldsymbol{\Sigma}_{t,l}\mathbf{1})^{-1}\left(y_{t+l} - \mathbf{1}'\boldsymbol{\Gamma}_{t,l}\boldsymbol{r}_{1:t}\right)^2 - \frac{1}{2}(\boldsymbol{x}_t - \boldsymbol{\mu}_x - \boldsymbol{\beta}\mathbf{1}'\boldsymbol{r}_t)'\boldsymbol{Q}^{-1}\right.$$

$$\left. (\boldsymbol{x}_t - \boldsymbol{\mu}_x - \boldsymbol{\beta}\mathbf{1}'\boldsymbol{r}_t) - \frac{1}{2}(\boldsymbol{r}_t - \boldsymbol{\Gamma}_{t-1,1}\boldsymbol{r}_{1:(t-1)})'\boldsymbol{\Sigma}_{t-1,1}^{-1}(\boldsymbol{r}_t - \boldsymbol{\Gamma}_{t-1,1}\boldsymbol{r}_{1:(t-1)})\right), \tag{A5}$$

whose right hand side is an exponential function of a quadratic form of y_{t+l} and $\boldsymbol{r}_t$. Hence, $y_{t+l}, \boldsymbol{r}_t \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}$ is multivariate normal. Consequently, the marginal distribution of $y_{t+l} \mid$

9

$\boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}$ is univariate normal. Moreover, the conditional expectation is

$$
\begin{aligned}
& E\left(y_{t+l} \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \\
=\ & E\left(E\left(y_{t+l} \mid \boldsymbol{r}_{1:t}\right) \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \\
=\ & E\left(\boldsymbol{\Gamma}_{t,l}(\boldsymbol{r}'_{1:(t-1)}, \boldsymbol{r}'_t)' \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}\right) \\
=\ & \boldsymbol{\Gamma}_{t,l}\left(\boldsymbol{r}'_{1:(t-1)}, E(\boldsymbol{r}_t \mid \boldsymbol{x}_t, \boldsymbol{r}_{1:(t-1)})'\right)',
\end{aligned}
$$

where $E(\boldsymbol{r}_t \mid \boldsymbol{x}_t, \boldsymbol{r}_{1:(t-1)})$ is linear in $\boldsymbol{x}_t$ and $\boldsymbol{r}_{1:(t-1)}$. Therefore, $y_{t+l} \mid \boldsymbol{x}_{t_0:t}, \boldsymbol{r}_{1:(t-1)}$ is univariate normal with mean linear in $\boldsymbol{x}_t$ and $\boldsymbol{r}_{1:(t-1)}$.

A7. Effect of Regularization

We now compare different types of regularization in fitting the coefficients of PRISM. The results are given in Table A4. Note that for L_2 regularization, we replaced all the L_1-norm penalties in (2.5) with L_2 norms (i.e., using the Ridge regression). We can see the clear advantages of adopting L_1 in error reduction, especially in farther-ahead forecasts. Unlike L_1 regularization, the L_2 penalty will not give zeros in the fitted coefficients. That is, those less relevant search terms will remain in the model with small, non-zero coefficients, which may lead to more variability in the prediction. However, with the L_1 penalty, the noise and redundant information from these search terms tend to be eliminated while only the most predictive search terms are selected to remain in the model. The results here show that the noise reduction by L_1 penalty is quite effective in improving the performance of PRISM.

Table A4: The performance of PRISM under different regularization over $2007 - 2016$. RMSE is measured relative to the respective error of the naive method. The boldface highlights the best performance for each forecasting horizon.

	real-time	forecast 1 week	forecast 2 week	forecast 3 week
PRISM	**0.498**	**0.492**	**0.453**	**0.467**
PRISM with L_2 regularization	0.537	0.553	0.534	0.544
naive	1	1	1	1

A8. The Length of Training Window

Here we conducted an empirical analysis on the choice of the rolling window length in training PRISM. With the limited, 5-year availability of weekly Google search data (more details in Section 2.1 and Supplementary Material A1), we varied the rolling window lengths between 2 and 4 years to keep the fitting and prediction within the same set of downloaded data. From Table A5, we see that the default 3-year window gives the leading performance. It appears the 3-year window length provides a good trade-off between the timeliness of short-term training data and the statistical efficiency from long-term, large-size data. The 3-year rolling window choice is also consistent with D'Amuri and Marcucci (2017). In addition, since Google data only starts in 2004, the choice of 3-year training window enables us to have forecasts from 2007 onward, which is

important for evaluating the performance of PRISM (and other methods) during the entire period of the financial crisis.

Table A5: The performance of PRISM under different rolling windows of training. RMSE is measured relative to the respective error of the naive method. The boldface highlights the best performance for each forecasting horizon.

	real-time	forecast 1 week	forecast 2 week	forecast 3 week
PRISM (default 3-year window)	**0.498**	**0.492**	**0.453**	**0.467**
PRISM with 2-year window	0.536	0.517	0.486	0.496
PRISM with 4-year window	0.499	0.496	0.459	0.468
naive	1	1	1	1

A9.   Normality of Residuals

Here we provide empirical evidence on the normality of the residuals for constructing the predictive intervals. Figure A3 shows the normal Q-Q plot of the fitted residuals by PRISM from 2008-2016. Except for a few points, the vast majority of the fitted residuals fall along with the normal distribution, which supports the construction of the predictive intervals by PRISM in Section 2.8 of the main text.
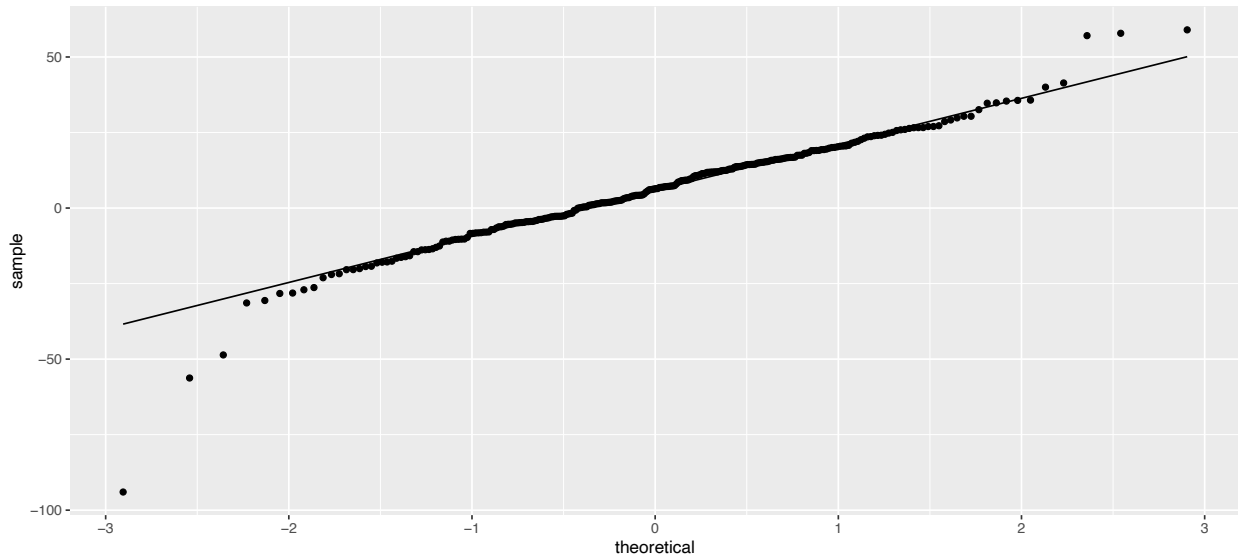

Figure A3: The normal Q-Q plot of fitted residuals by PRISM in $2008 - 2016$.

A10.   Fitted Coefficients of PRISM

We report the dynamically fitted coefficients by PRISM from each week's prediction in 2008-2016 in Figure A4. Focusing on the coefficients of Google search terms (the bottom section of the

heatmap), we can see the sparsity of coefficients from the L_1 penalty. On average, 6.20 out of the 25 search terms are selected and included in the each week's model by PRISM during 2008-2016. Note that all 25 terms have been included at some point. In addition, we can also see different search terms come in and out of the model over time, indicating the dynamic movement for each term in its contribution to the final forecasting.
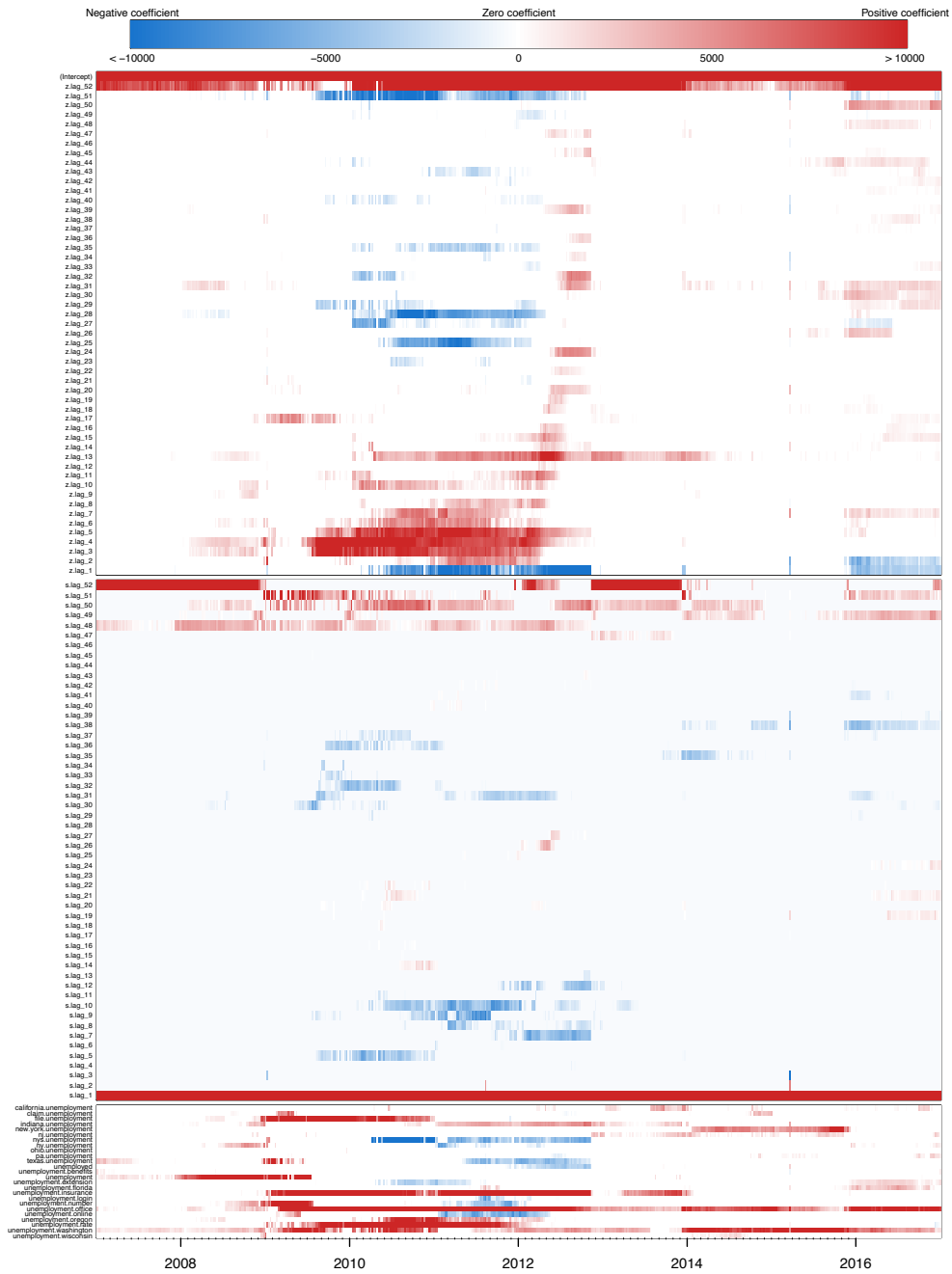
Figure A4: The heatmap illustrating the dynamically fitted coefficients by PRISM for each week of forecasting in 2008 − 2016. Red color represents positive coefficients, blue color represents negative coefficients, and white color represents zero.