DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

Daniel Borup, David E. Rapach and
Erik Christian Montes Schütte

CREATES Research Paper 2021-02

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

Daniel Borup
Aarhus University, CREATES, and DFI
dborup@econ.au.dk

David E. Rapach*
Washington University in St. Louis and Saint Louis University
david.rapach@wustl.edu

Erik Christian Montes Schütte
Aarhus University, CREATES, and DFI
christianms@econ.au.dk

January 7, 2021

*Corresponding author. Send correspondence to David Rapach, Olin School of Business, Washington University in St. Louis, St. Louis, MO 63130; email: david.rapach@wustl.edu. We are grateful to Scott Brave, Marco Colagrossi, Philippe Goulet Coulombe, Francesco D'Amuri, Andrea Geraci, Edward Kong, Michael McCracken, Daniel Prinz, Dalibor Stevanovic, and Nicolas Woloszko, as well as conference and seminar participants at the 2020 Banca d'Italia and Federal Reserve Board Joint Conference on Nontraditional Data and Statistical Learning with Applications to Macroeconomics, Competence Centre on Microeconomic Evaluation Seminar Series at the European Commission Joint Research Centre, Dale T. Mortensen COVID-19 Workshop, 2020 International Conference on Computational and Financial Econometrics, Aarhus University, and Instituto Tecnológico Autónomo de México (ITAM), for valuable comments. Borup and Schütte are grateful for research support from the Center for Research in Econometric Analysis of Time Series (CREATES) and Danish Finance Institute (DFI). The authors have no potential conflicts of interest to disclose.

Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data

Abstract

We generate a sequence of now- and backcasts of weekly unemployment insurance initial claims (UI) based on a rich trove of daily Google Trends (GT) search-volume data for terms related to *unemployment*. To harness the information in a high-dimensional set of daily GT terms, we estimate predictive models using machine-learning techniques in a mixed-frequency framework. In a simulated out-of-sample exercise, now- and backcasts of weekly UI that incorporate the information in the daily GT terms substantially outperform models that ignore the information. The relevance of GT terms for predicting UI is strongly linked to the COVID-19 crisis.

JEL classifications: C45, C53, C55, E24, E27, J65

Keywords: Unemployment insurance, Internet search, Mixed-frequency data, Penalized regression, Neural network, Variable importance

1. Introduction

The COVID-19 crisis has created economic upheaval in the United States, including historically unprecedented levels of unemployment insurance initial claims (UI). After a national emergency was declared on March 13, 2020 and closures of non-essential retail establishments were ordered in many parts of the country, UI spiked in late March, reaching a (seasonally adjusted) record of 6,867,000 for the week ending March 28, 2020. By comparison, the peak in UI during the Great Recession was "only" 665,000 (for the week ending March 28, 2009). While UI has subsequently declined, it remains at elevated levels. Because it provides important information about the US labor market and is reported at the weekly frequency, UI has become perhaps the most closely watched economic variable during the COVID-19 crisis. Reflecting its relevance and timeliness, Lewis, Mertens, and Stock (2020) include UI in their recently developed weekly economic indicator for the United States.

In this paper, we use a rich trove of daily internet search-volume data from Google Trends (GT) to predict UI, with an eye toward improving prediction during the COVID-19 crisis. Because we cannot know a priori the most relevant GT search terms for predicting UI, we employ a high-dimensional set of GT terms related to *unemployment* that reflects individuals' searches for information about filing for unemployment benefits when they become (or anticipate becoming) unemployed. We then rely on machine-learning techniques to harness the relevant information in the terms. Specifically, we use daily data for 103 *unemployment*-related GT terms for the most recent seven days to generate a sequence of now- and backcasts of a given week's (Sunday through Saturday) UI, in anticipation of the figure's release by the Department of Labor on Thursday of the following week. The sequence of now- and backcasts incorporates the most recent daily GT data as they become available, which allows us to investigate the "term structure" of the flow of information with respect to predictive accuracy. The sequence of now- and backcasts of week-t UI are made from ten days to one day before the UI release on Thursday of week $t + 1$.

Each of our predictive models relates UI to its first or second lag—in recognition of the serial correlation in UI—as well as seven days of GT data.[1] Each model thus contains $7 \times 103 + 1 = 722$ predictors (or inputs), so that ours is a high-dimensional setting. We begin with a linear specification for the predictive models underpinning the now- and backcasts. We estimate the linear predictive models via the least absolute shrinkage and selection operator (LASSO, Tibshirani 1996) and elastic net (ENet, Zou and Hastie 2005). The LASSO and ENet are popular machine-learning devices, which improve prediction in high-dimensional settings by including a penalty term in the objective function for estimating the model's parameters. Intuitively, the penalty term works to shrink the parameters toward zero, thereby helping to prevent overfitting. Because their penalty terms include an ℓ_1 component, the LASSO and ENet permit shrinkage to zero; hence, they facilitate model interpretation by performing variable selection.

To allow for more complex, nonlinear predictive relationships, we also employ artificial neural networks (ANNs) to generate now- and backcasts of UI based on the 722 predictors. ANNs contain one or more hidden layers, each of which contains multiple neurons that transmit predictive signals through the network. Under a reasonable set of restrictions, a single-layer ANN with a sufficient number of neurons can approximate any smooth function (e.g., Cybenko 1989; Funahashi 1989; Hornik, Stinchcombe, and White 1989; Hornik 1991; Barron 1994). Because ANNs with multiple hidden layers are often used in practice, we consider ANNs with one to three hidden layers (NN1, NN2, and NN3, respectively).[2] We fit the ANNs using the recently developed Adam stochastic gradient descent (SGD) algorithm (Kingma and Ba 2015).

Our use of mixed-frequency data in the predictive models is a version of the unrestricted mixed-data sampling (U-MIDAS) approach of Foroni, Marcellino, and Schumacher (2015). The restricted MIDAS approach (Ghysels, Santa-Clara, and Valkanov 2005) imposes a lag-

[1] The inclusion of the first or second lag of UI is determined by the timing of the UI release, as explained in Section 3.

[2] ANNs with one or two (three or more) hidden layers are typically referred to as shallow (deep) neural networks.

polynomial structure on the higher-frequency data. Instead of somewhat arbitrarily imposing a lag-polynomial structure, we harness machine-learning techniques to fit the weights for the individual variables in the predictive models in our high-dimensional setting. The U-MIDAS framework also allows us to analyze how the flow of information affects the accuracy of the sequence of now- and backcasts, as each successive element in the sequence includes a more recent day of GT data. In essence, we combine two branches of the economic forecasting literature, one that applies machine-learning tools (e.g., Diebold and Shin 2019; Kotchoni, Leroux, and Stevanovic 2019; Borup and Schütte forthcoming; Medeiros et al. forthcoming) and one that employs mixed-frequency data (e.g., Clements and Galvão 2008; Foroni and Marcellino 2014; Brave, Butters, and Justiniano 2019). To our knowledge, our study is the first to use high-dimensional, mixed-frequency data together with the U-MIDAS approach and machine-learning techniques to generate predictions, so that we make a methodological contribution along this dimension.

We find that the information in our high-dimensional set of daily GT terms is indeed useful for now- and backcasting weekly UI, even up to ten days before its release date. This finding holds for both non-seasonally and seasonally adjusted UI. Specifically, the predictions for models that include GT terms generate substantial improvements in root mean squared error (RMSE) vis-à-vis an autoregressive (AR) benchmark model. For an out-of-sample period spanning the first week of January 2015 through the first week of August 2020, all of the now- and backcasts based on the daily GT trends deliver a lower RMSE than the AR benchmark. Many of the reductions in RMSE are quite sizable, reaching as high as 63%. The improvement in predictive accuracy offered by the daily GT terms is dramatic during the advent of the COVID-19 crisis.

The now- and backcasts based on both linear models and ANNs perform well overall. Predictions based on ANNs typically perform better than those based on linear models for non-seasonally adjusted UI, while the converse holds for seasonally adjusted UI. We also

consider three ensembles for a given now- or backcast of UI.[3] The first takes the average of the LASSO and ENet predictions, the second is the average of the predictions generated by the three ANNs, and the last is the average of all five predictions (LASSO, ENet, NN1, NN2, and NN3). The ensemble based on all of the individual predictions works quite well, producing a reduction in RMSE vis-à-vis the AR benchmark that is nearly as large as that of the best individual model (which we cannot know a priori).

The term structure of the information flow reveals a pronounced increase in predictive accuracy as more timely daily GT data are incorporated into the sequence of now- and backcasts. For the first nowcast, which is made ten days before the UI release, the GT terms improve the RMSE vis-à-vis the AR benchmark by approximately 25% (15%) on average for non-seasonally (seasonally) adjusted UI. For the backcast made three days before the UI release, which incorporates GT data for seven more recently available days, the improvement in RMSE is around 55% (45%) on average for non-seasonally (seasonally) adjusted UI. The best overall performance obtains when there is full overlap between the seven days of GT terms and the UI week, corresponding to the backcast made three days before the UI release.

We also test whether the predictive accuracy of the models that incorporate the GT terms relative to that of the AR benchmark varies with social conditions related to the COVID-19 crisis. Using four measures of social conditions (UI minus its pre-crisis average, change in COVID-19 deaths, government response stringency index, and workplace closure stringency sub-index), we find that the relative performance of the models that incorporate the GT terms significantly improves as social conditions worsen. This provides further evidence that the predictive information in the GT terms is linked to the COVID-19 crisis.

To look inside the "black box" of the fitted ANNs, we compute variable-importance measures (Greenwell, Boehmke, and McCarthy 2018) and partial-dependence plots (PDPs, Friedman 2001) for the individual predictors. The variable-importance measures allow us to see which predictors are the most relevant in the fitted ANNs, as well as the linear

[3]Ensembles are often refereed to as combinations in the econometrics literature; see Timmermann (2006) for a survey of forecast combinations.

models fitted via the LASSO and ENet. Prior to the COVID-19 crisis, the lag of UI is the most important predictor in all of the fitted models. With the advent of the crisis, the situation changes markedly, as GT terms related to the application process for UI benefits (e.g., *apply for unemployment benefits*, *file for unemployment*, and *unemployment office*) become substantially more important. The PDPs allow us to analyze the strength of the nonlinearities in the fitted ANNs. For non-seasonally adjusted UI, there is evidence of nonlinearities in the predictive relationships between the GT terms and UI, with the effect of the GT terms becoming stronger as the terms take on more extreme values in connection with the COVID-19 crisis.

We contribute to an emerging literature that uses internet search-volume data to predict labor market variables. For example, D'Amuri and Marcucci (2017) show that GT search volume for terms including *jobs* improves US unemployment rate prediction, while Niesert et al. (2020) find that an array of GT terms are useful for predicting unemployment rates in a collection of developed countries. Borup and Schütte (forthcoming) use a large number of GT terms and machine-learning tools to improve US employment growth prediction. In contrast to these studies, which predict variables available at the monthly frequency, our target is weekly UI—which has become perhaps the most closely watched variable since the advent of the COVID-19 crisis—and we use a mixed-frequency approach. Some recent studies use GT terms to predict UI during the crisis (Aaronson et al. 2020; Goldsmith-Pinkham and Sojourner 2020; Larson and Sinclair 2020). Unlike the present paper, these studies consider only a small number of GT terms and do not utilize machine-learning methods.[4]

The rest of the paper is organized as follows. Section 2 describes the data, while Section 3 explains the information flow for the sequence of now- and backcasts. Section 4 specifies the predictive models and outlines their estimation. Section 5 reports results for the out-of-

[4]Aaronson et al. (2020) use an event-study design based on the sensitivity of UI to hurricanes to predict UI during the start of the COVID-19 crisis using GT terms, while Goldsmith-Pinkham and Sojourner (2020) predict UI during the start of the crisis using the GT term *file for unemployment*. Larson and Sinclair (2020) use a small number of GT terms in panel regressions to nowcast UI across US states; in contrast to our study, nowcasts based on the GT terms fail to outperform those based on an AR benchmark. Choi and Varian (2012) use a small number of GT terms to predict UI through mid 2011.

sample exercise. Section 6 interprets the fitted models via variable-importance measures and PDPs. Section 7 concludes.

2. Data

This section describes the data, which span the first week of January 2005 to the first week of August 2020.

2.1. Unemployment Insurance Initial Claims

Our target variable is UI for the United States. UI is available at the weekly frequency, corresponding to initial claims for Sunday through Saturday. Each Thursday morning at 8:30 EST, the Department of Labor releases the UI figure for the previous week. We take this publication lag into account when computing our predictions.[5] As detailed in Section 3, we are careful in tracking the information flow, so that we only use information available at the time of prediction formation.

The choice between targeting non-seasonally or seasonally adjusted data has been the subject of recent debate during the COVID-19 crisis (e.g., Rinz 2020). The issue is whether the conventional multiplicative seasonal-adjustment process overstates the actual seasonality in the data during the crisis, when UI reached historically unprecedented levels.[6] To address this issue, we generate predictions for both non-seasonally and seasonally adjusted UI.

2.2. Google Trends

Daily internet search-volume data are obtained from GT, which provides an index of the proportion of queries for a specific search term within a geographical area. The index is

[5]Compared to other macroeconomic variables (e.g., gross domestic product and consumption), UI data are subject to relatively minor revisions, as UI is based on government administrative data (rather than surveys). UI is typically revised only once during the following week. We accommodate this via vintage data when computing our predictions.

[6]Indeed, the Department of Labor recently switched to an additive seasonal-adjustment process for UI (Davidson 2020). The change was only made going forward, so that historical data will not be revised.

released with a maximum delay of 36 hours. This delay is the result of Google filtering irregular search activity, such as automated searches or queries that may be associated with attempts to spam search; see "FAQ about Google Trends data."

We construct a high-dimensional set of predictors based on daily GT terms. Starting with the source term *unemployment*, we use Google Keyword Planner, which provides the most relevant terms to include in a webpage to increase web traffic, to obtain the following top 15 keywords associated with this term: (1) *unemployment*, (2) *unemployment benefits*, (3) *unemployment office*, (4) *unemployment insurance*, (5) *file for unemployment*, (6) *apply for unemployment*, (7) *unemployment claim*, (8) *how to file for unemployment*, (9) *ui online*, (10) *unemployment application*, (11) *unemployment weekly claim*, (12) *unemployment compensation*, (13) *unemployment number*, (14) *unemployment online*, (15) *employment insurance*. These "primitive terms" appear quite plausible, as they are associated with the actions of a person who becomes unemployed. Our out-of-sample period begins in 2015 in Section 5, so that, to avoid look-ahead bias, the set of primitive terms is based on GT data through the end of 2014.

We expand each of the primitive terms via a GT feature that provides a list of 25 related terms, again based on GT data through the end of 2014. We use the *top* category of related terms (instead of the *rising* category). This step adds terms that are specific to individual US states (e.g., *ny unemployment benefits*, *unemployment benefits california*); semantically related to the primitive terms (e.g., *how to apply for unemployment benefits*, *unemployment phone number*, *filing unemployment online*, *state unemployment office*); closely related to unemployment, such as health care coverage and tax policies (e.g., *unemployment health insurance*, *unemployment insurance taxes*); and narrowly defined (e.g., *edd online*, which refers to the Employment Development Department, through which unemployment insurance benefits can be applied for in California). After excluding duplicates, this produces a total of 270 keywords.

After removing low-volume queries (defined as series with less than 95% non-zero values), we have 103 unique terms at the daily frequency. In the context of predicting employment growth, Borup and Schütte (forthcoming) find that minor variations in the wording of queries (like adding or removing an s for a plural or singular version of a word) can have a notable influence on their predictive power. We cannot know a priori which specific terms or variations are the most relevant for predicting UI, so that we include a large number of related terms and rely on supervised machine-learning methods to place greater weight on those that are deemed the most relevant. Terms that are specific to individual US states capture idiosyncrasies for each state, which is useful if, say, New York is suddenly the main driver of unemployment claims. This is relevant, given the heterogeneity in the spreading of COVID-19 and political responses across different parts of the United States. Table 1 lists the 103 unique GT terms for our analysis.

GT only allows for the downloading of daily data in blocks that do not cover the full sample period, so that we concatenate data from each download to construct complete time series. The downloaded data for each GT term are scaled to have a value of 100 for the day with the highest volume. We thus need to adjust the levels of each downloaded block of data to chain together series that are comparable over time. To accomplish this, we download seven-month blocks of data, with a month of overlap. For each GT term, we compute the average daily value for the current and preceding blocks for the overlapping month. We then use the ratio of the two averages to adjust the levels for all preceding blocks.[7]

Each block of downloaded GT data covering a particular period is based on a randomized sample (about 1%) of total search queries during the period. The values for the block corresponding to the period thus change according to the time and IP address of the request

[7] A simple example illustrates the basic idea. Suppose that the first set of downloaded data for an arbitrary series is 90 and 99 for periods 1 and 2, respectively; the next set of downloaded data is 85 and 76 for periods 2 and 3, respectively. We take the ratio of the values for period 2, $85/99 = 0.86$; we then use the ratio to adjust the period 1 value to $90 \times 0.86 = 77.27$, which gives us a comparable series of 77.27, 85, and 76 for periods 1, 2, and 3, respectively.

to download the data. To reduce sampling error, we make ten requests for a particular period and take the average of the values over the ten downloaded blocks.

Finally, we seasonally adjust each of the GT terms using the popular STL filtering procedure (Cleveland et al. 1990). To avoid look-ahead bias, we recursively seasonally adjust the GT terms using data available at the time of prediction formation.

Figure 1 depicts UI, along with two selected GT terms (*file for unemployment* and *unemployment office*), for the first week of January 2020 to the first week of August 2020.[8] The time stamp on the horizontal axis indicates the UI release. The upper (lower) panels display results for non-seasonally (seasonally) adjusted UI. For both the non-seasonally and seasonally adjusted cases, UI exhibits a dramatic increase for the March 26 release, corresponding to the week ending March 21, followed by another sharp increase in the next week, leading to an historical high of approximately 6.2 (6.9) million on a non-seasonally (seasonally) adjusted basis for the week ending March 28. UI then decreases gradually, although it remains quite elevated from an historical perspective.

The two GT terms in Figure 1 appear to track UI quite well. Specifically, the terms start to increase markedly in the weeks around the sharp increase in UI, and they follow the subsequent downward trajectory fairly closely. Figure 1 suggests that GT terms are relevant for predicting UI. This is economically intuitive, as individuals are likely to search for information about filing for unemployment benefits when they become (or anticipate becoming) unemployed. Such searches leave a footprint in the search volume of relevant queries, which we harness to predict UI.

3. Information Flow

Table 2 explains the flow of information for generating our sequence of predictions. In terms of notation, we denote the days comprising week t by Sunday_t, Monday_t, ... , Saturday_t. The Department of Labor releases the week-$(t-1)$ UI figure (UI_{t-1}) on Thursday_t. Since

[8] These are two of the most important terms for predicting UI in Section 6.

GT data are released with a maximum 36-hour delay, we employ a conservative two-day lag, so that search-volume data for queries for, say, Saturday$_t$ are available on Monday$_{t+1}$. When generating each prediction, we use the seven most recently available daily observations for each of the 103 GT terms.

We begin with a prediction of UI$_t$ formed on Monday$_t$, which corresponds to a nowcast of UI$_t$. After accounting for the maximum 36-hour reporting lag, the seven most recently available daily observations for the GT terms cover Sunday$_{t-1}$ through Saturday$_{t-1}$. We compute the nowcast by first using historical data available at the time of prediction formation to estimate one of the predictive models described in Section 4, which relates UI for a given week to GT terms for the seven days in the previous week, as well as the second lag of UI. The UI lag accounts for the strong autocorrelation in UI. We use the second lag, because, as indicated in the last column of Table 2, the most recent UI observation available for computing the nowcast of UI$_t$ is for week $t-2$ (due to the reporting lag for UI). We then plug the values for the GT terms for Sunday$_{t-1}$ through Saturday$_{t-1}$ and most recent UI observation (UI$_{t-2}$) into the fitted model to generate the nowcast of UI$_t$.

Next is a prediction of UI$_t$ formed on Tuesday$_t$, which again corresponds to a nowcast. Since an additional day of GT data is available, this nowcast is based on terms for Monday$_{t-1}$ to Sunday$_t$, so that there is now a one-day overlap between the GT terms and UI$_t$; see the fourth column of Table 2. To compute the nowcast, we first fit the predictive model, which relates UI in a given week to GT terms for Sunday of that week and Monday through Saturday of the previous week (as well as the second lag of UI). We then plug the values for the GT terms for Monday$_{t-1}$ to Sunday$_t$ (and UI$_{t-2}$) into the fitted model. We proceed analogously to compute the nowcast of UI$_t$ formed on Wednesday$_t$, which is characterized by a two-day overlap between the GT terms and UI$_t$.

The next three nowcasts in Table 2 are formed on Thursday$_t$, Friday$_t$, and Saturday$_t$. In addition to incorporating GT data through Tuesday$_t$, Wednesday$_t$, and Thursday$_t$, respectively, the latest available UI release allows us to use UI for week $t-1$. We thus use the

first (instead of the second) lag of UI in the predictive model. Observe that as we move from Thursday$_t$ to Saturday$_t$ when forming the nowcasts, we go from a three- to a five-day overlap between the available GT terms and UI$_t$. Otherwise, we compute the nowcasts in the same manner as the first three nowcasts in Table 2.

The remaining predictions of UI$_t$, formed on Sunday$_{t+1}$ through Wednesday$_{t+1}$, constitute backcasts. The backcast formed on Monday$_{t+1}$ employs the maximum overlap of seven days between the available GT terms and UI$_t$. The backcast formed on Tuesday$_{t+1}$ (Wednesday$_{t+1}$) uses GT data for six (five) days from week t and one (two) day(s) from week $t+1$.

The sequence of predictions in Table 2 allows us to investigate the term structure of the information flow with respect to predicting UI. As we proceed from the nowcast formed on Monday$_t$ to the backcast formed on Monday$_{t+1}$, the degree of overlap between the days used to predict UI$_t$ increases. For the final two backcasts in Table 2, we include GT terms from the first one or two days of week $t+1$ when predicting UI$_t$. We are interested in how the availability of more recent daily GT data affects the accuracy of the now- and backcasts.

4. Predictive Models

The general form of a predictive model is given by

$$\text{UI}_t = f^{(j)}\Big(\text{UI}_{t-1(2)}, \boldsymbol{g}_t^{(j)}; \boldsymbol{\theta}^{(j)}\Big), \tag{4.1}$$

where $\boldsymbol{\theta}^{(j)}$ is a vector of model parameters specific to $f^{(j)}$;

$$\underbrace{\boldsymbol{g}_t^{(j)}}_{7K\times 1} = \begin{bmatrix} \boldsymbol{g}'_{t-j/7} & \boldsymbol{g}'_{t-(j+1)/7} & \cdots & \boldsymbol{g}'_{t-(j+6)/7} \end{bmatrix}'; \tag{4.2}$$

$\boldsymbol{g}_{t-i/7}$ is a $K \times 1$ vector of GT terms for the $(7-i)$th day of week t for $i = 0, \ldots, 6$; and $K = 103$.[9] The fifth column of Table 2 provides the value of j for each of the now- and backcasts; the data overlap in the fourth column is given by $7 - |j|$. Based on data availability for UI, we use UI_{t-2} for the first three nowcasts and UI_{t-1} for the remaining now- and backcasts in Table 2, explaining the subscript notation for the AR term in Equation Equation (4.1).[10]

We begin with a linear specification for the predictive model:

$$\text{UI}_t = \alpha^{(j)} + \beta_{\text{AR}}^{(j)} \text{UI}_{t-1(2)} + \boldsymbol{\beta}_g^{(j)\prime} \boldsymbol{g}_t^{(j)} + \varepsilon_t^{(j)}, \tag{4.3}$$

where $\boldsymbol{\beta}_g^{(j)}$ is a $7K \times 1$ vector of slope coefficients for the daily GT terms and $\varepsilon_t^{(j)}$ is zero-mean error term. In terms of Equation (4.1), the vector of model parameters is given by

$$\boldsymbol{\theta}^{(j)} = \begin{bmatrix} \alpha^{(j)} & \beta_{\text{AR}}^{(j)} & \boldsymbol{\beta}_g^{(j)\prime} \end{bmatrix}'. \tag{4.4}$$

There are $7 \times 103 + 1 = 722$ regressors in Equation (4.3). In our high-dimensional setting, we use the LASSO and ENet to estimate $\boldsymbol{\theta}^{(j)}$ in Equation (4.3) when generating the now- and backcasts.[11]

Equation (4.3) can be viewed as a U-MIDAS model (Foroni, Marcellino, and Schumacher 2015), as it allows each of the higher-frequency predictors in $\boldsymbol{g}_t^{(j)}$ to have its own coefficient. A restricted MIDAS specification imposes a lag-polynomial structure on the daily observations. We use a U-MIDAS approach for two reasons. First, the daily observations only naturally align with the calendar week when $j = 0$ or $j = 7$ (see Table 2). Second, we employ machine-

[9] The vector of GT terms for each day of week t is as follows: Sunday, $\boldsymbol{g}_{t-6/7}$; Monday, $\boldsymbol{g}_{t-5/7}$; Tuesday, $\boldsymbol{g}_{t-4/7}$; Wednesday, $\boldsymbol{g}_{t-3/7}$; Thursday, $\boldsymbol{g}_{t-2/7}$; Friday, $\boldsymbol{g}_{t-1/7}$; Saturday, $\boldsymbol{g}_t$.

[10] Including additional lags for UI in Equation (4.1) has little effect on the results, so that a single lag appears sufficient for capturing the autocorrelation in UI.

[11] When we compute simulated out-of-sample now- and backcasts starting in the first week of January 2015, the number of weekly UI observations available for fitting the predictive model is 520, which is less than the number of predictors (722), so that the conventional ordinary least squares estimator fails.

learning methods that allow us to flexibly estimate the weights—rather than somewhat arbitrarily imposing a lag-polynomial structure—while guarding against overfitting.[12]

4.1. Penalized Regression

The LASSO (Tibshirani 1996) is a machine-learning device based on penalized regression. It alleviates overfitting by augmenting the objective function for estimating $\boldsymbol{\theta}^{(j)}$ in Equation (4.3) with an ℓ_1 penalty term:

$$\underset{\boldsymbol{\theta}^{(j)}\in\mathbb{R}^{7K+2}}{\arg\min}\ \frac{1}{2\mathcal{T}}\left\{\sum_{t=1}^{\mathcal{T}}\left[\mathrm{UI}_t-\left(\alpha^{(j)}+\beta_{\mathrm{AR}}^{(j)}\mathrm{UI}_{t-1(2)}+\boldsymbol{\beta}_g^{(j)\prime}\boldsymbol{g}_t^{(j)}\right)\right]^2\right\}+\lambda\|\boldsymbol{\beta}^{(j)}\|_1, \tag{4.5}$$

where

$$\boldsymbol{\beta}^{(j)}=\begin{bmatrix}\beta_{\mathrm{AR}}^{(j)} & \boldsymbol{\beta}_g^{(j)\prime}\end{bmatrix}', \tag{4.6}$$

$\mathcal{T}$ is the number of weekly UI observations used to fit the model, $\|\cdot\|_1$ is the ℓ_1 norm, and $\lambda\geq 0$ is a regularization parameter for controlling the degree of shrinkage. Unlike the ℓ_2 penalty in ridge regression (Hoerl and Kennard 1970), the ℓ_1 penalty in Equation (4.5) permits shrinkage to zero (for sufficiently large λ), so that the LASSO performs variable selection.

Although the LASSO is effective at selecting relevant predictors in certain environments (e.g., Zhang and Huang 2008; Bickel, Ritov, and Tsybakov 2009; Meinshausen and Yu 2009), it tends to arbitrarily select one predictor from a group of highly correlated predictors. The ENet (Zou and Hastie 2005) is a refinement of the LASSO, which mitigates this tendency by including both ℓ_1 (LASSO) and ℓ_2 (ridge) components in the penalty term for the objective

[12]Foroni, Marcellino, and Schumacher (2015) find that a "small" difference in sampling frequency between the higher- and lower-frequency variables (as in our application) favors the U-MIDAS over the restricted MIDAS approach.

function:

$$\underset{\boldsymbol{\theta}^{(j)}\in\mathbb{R}^{7K+2}}{\arg\min}\ \frac{1}{2\mathcal{T}}\left\{\sum_{t=1}^{\mathcal{T}}\left[\mathrm{UI}_t-\left(\alpha^{(j)}+\beta_{\mathrm{AR}}^{(j)}\mathrm{UI}_{t-1(2)}+\boldsymbol{\beta}_g^{(j)\prime}\boldsymbol{g}_t^{(j)}\right)\right]^2\right\}+\lambda P_\delta\left(\boldsymbol{\beta}^{(j)}\right), \tag{4.7}$$

where

$$P_\delta\left(\boldsymbol{\beta}^{(j)}\right)=0.5(1-\delta)\|\boldsymbol{\beta}^{(j)}\|_2^2+\delta\|\boldsymbol{\beta}^{(j)}\|_1, \tag{4.8}$$

$\|\cdot\|_2$ is the ℓ_2 norm, and $0\leq\delta\leq 1$ is a blending parameter for the ℓ_1 and ℓ_2 components of the penalty term. When $\delta=1$, $P_\delta=\|\boldsymbol{\beta}^{(j)}\|_1$ in Equation (4.8), so that the ENet reduces to the LASSO. We follow the recommendation of Hastie and Qian (2016) and set $\delta=0.5$.[13]

After estimating $\boldsymbol{\theta}^{(j)}$ in Equation (4.3) via the LASSO or ENet using data available at the time of prediction formation, we plug the most recently available UI observation and seven most recently available daily observations for each of the 103 GT terms into the fitted model to generate a given now- or backcast in Table 2.

4.2. Artificial Neural Networks

We consider feedforward ANNs, which have proven useful for prediction in numerous domains. An ANN architecture is comprised of multiple layers. The first, the input layer, is the set of predictors, which we denote by $x_1,\ldots,x_{P_0}$. One or more hidden layers follow. Each hidden layer l contains P_l neurons, each of which takes signals from the neurons in the previous hidden layer to generate a subsequent signal:

$$h_m^{(l)}=g\left(w_{m,0}^{(l)}+\sum_{j=1}^{P_{l-1}}w_{m,j}^{(l)}h_j^{(l-1)}\right)\ \text{for}\ m=1,\ldots,P_l;\ l=1,\ldots,L, \tag{4.9}$$

[13]To better guard against overfitting, we tune the regularization parameter λ for the LASSO and ENet in Equations (4.5) and (4.7), respectively, via the extended regularization information criterion (Hui, Warton, and Foster 2015), which is a refinement of the Bayesian information criterion (Schwarz 1978).

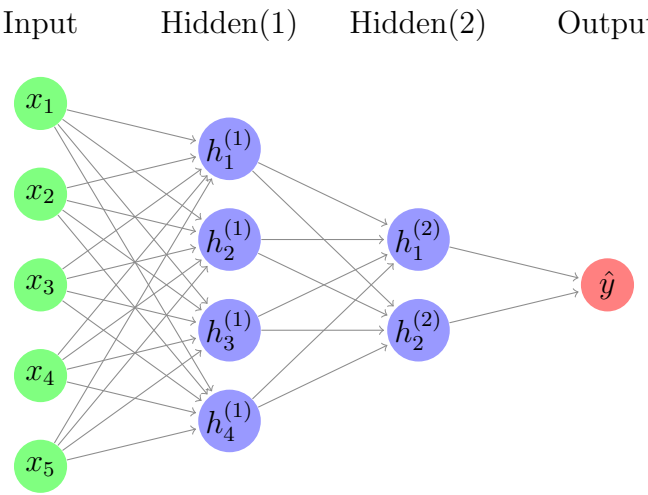
where $h_m^{(l)}$ is the signal corresponding to the mth neuron in the lth hidden layer;[14] $w_{m,0}^{(l)}, w_{m,1}^{(l)}, \ldots, w_{m,P_{l-1}}^{(l)}$ are weights; and $g(\cdot)$ is an activation function. The final layer is the output layer, which translates the signals from the last hidden layer into a prediction:

$$\hat{y} = w_0^{(L+1)} + \sum_{j=1}^{P_L} w_j^{(L+1)} h_j^{(L)}, \tag{4.10}$$

where $\hat{y}$ denotes the prediction of the target variable. For the activation function, we use the popular rectified linear unit (ReLU) function:

$$g(x) = \begin{cases} 0 & \text{if } x < 0, \\ x & \text{otherwise.} \end{cases} \tag{4.11}$$

In response to a sufficiently strong signal, Equation (4.11) activates a neuronal connection and relays the signal forward through the network. To illustrate the basic structure of an ANN, the following diagram portrays a feedforward ANN consisting of five inputs and two hidden layers with four and two neurons, respectively:



The myriad of interactions among the inputs and neurons in the network allows for complex nonlinear predictive relationships.

[14] For the first hidden layer, $h_j^{(0)} = x_j$ for $j = 1, \ldots, P_0$.

Theoretically, a single hidden layer with enough nodes is sufficient for approximating any smooth function (e.g., Cybenko 1989; Funahashi 1989; Hornik, Stinchcombe, and White 1989; Hornik 1991; Barron 1994). Nevertheless, ANNs with multiple hidden layers are often used in practice (e.g., Rolnick and Tegmark 2018). We consider ANNs with one, two, and three hidden layers (NN1, NN2, and NN3, respectively). We include 104 neurons in the single hidden layer for NN1, corresponding to the number of GT terms (103) for the different days plus the lag of UI. Based on the "pyramid" strategy (Masters 1993), NN2 (NN3) contains 104 and ten (104, ten, and three) neurons in its first and second (first, second, and third) hidden layers, respectively.

Fitting (or training) an ANN requires estimating the weights. We train the ANNs by minimizing an objective function based on the MSE for the training sample, which we augment with an ℓ_1 penalty term to help guard against overfitting. We use the recently developed Adam SGD algorithm (Kingma and Ba 2015) to train the ANNs, which we implement in `Python` using the `keras` package. In the algorithm, we set the number of epochs to 700 and batch size to 32 (the `keras` default).[15]

4.3. Ensembles

We also consider ensemble predictions, which are popular in machine learning. In recognition of model uncertainty, instead of relying on a prediction based on a single model, we take an average of the predictions generated by multiple models. For each of the now- and backcasts, we construct three ensembles. The first (Ensemble-Linear) is an average of the predictions based on the linear models fitted via the LASSO and ENet. The second (Ensemble-ANN) is an average of the predictions based on the fitted NN1, NN2, and NN3 models. The final ensemble (Ensemble-All) is an average of the predictions based on the linear models fitted via the LASSO and ENet and fitted NN1, NN2, and NN3 models.

[15]To reduce the influence of starting values for the random-number generator in the algorithm, we compute an ensemble prediction by training a given model ten different times with a different seed each time and taking the average of the ten predictions.

5. Out-of-Sample Results

We generate simulated out-of-sample now- and backcasts for UI for the first of week of January 2015 through the first week of August 2020. Simulating the situation of a forecaster in real time, we proceed as follows using a rolling-window approach. We first use UI and GT data available for the first week of January 2005 (when enough GT terms have sufficient volume) through the last week of December 2014 to fit the predictive models. We then plug the most recent UI value and relevant GT values into the fitted models to generate the now- and backcasts for UI for the first week of January 2015. Next, we refit the predictive models using UI and GT data available for the second week of January 2005 through the first week of January 2015 and plug the most recent UI and relevant GT values into the fitted models to compute the UI now- and backcasts for the second week of January 2015. We continue in this manner through the end of the out-of-sample period. We reiterate that the simulated now- and backcasts only use information available at the time of prediction formation, as described in Section 3.

Although it substantially increases computational cost, it is important to refit the models each week as new data become available. In particular, as we discuss in Section 6, the GT terms play a relatively limited role in models fitted using data that exclude the COVID-19 crisis, while they become substantively more important when the training sample includes data from the crisis. By refitting the models each week, we are able to capture the growing importance of the GT terms in a more timely manner.

An AR model based on the first or second lag of UI is used to generate the benchmark now- and backcasts, where we again fit the AR models using a rolling-window approach.[16] An AR model is a standard benchmark in the economic forecasting literature. Due to the significant autocorrelation in UI, it is a relevant benchmark in our application, and it

[16]Based on the information flow (see Section 3), the AR benchmark model is given by $UI_t = \rho_0 + \rho_2 UI_{t-2} + \varepsilon_t$ for the nowcasts formed on Monday_t through Wendesday_t; it is given by $UI_t = \rho_0 + \rho_1 UI_{t-1} + \varepsilon_t$ for the now- and backcasts formed on Thursday_t through Wendesday_{t+1}. We estimate the AR models via ordinary least squares.

performs substantially better than a naïve model that ignores the autocorrelation in UI and simply uses the rolling mean to predict UI.

5.1. Predictive Accuracy

Panel A (B) of Table 3 reports results for non-seasonally (seasonally) adjusted UI.[17] The third column reports the RMSE for the AR benchmark for each of the now- and backcasts. The MSE is defined as

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}\left(\text{UI}_t - \widehat{\text{UI}}_t\right)^2, \tag{5.1}$$

where $\widehat{\text{UI}}_t$ generically denotes a prediction of UI_t and T is the number of observations available for analyzing the predictions. RMSE is the square root of Equation (5.1).

For the nowcasts formed on Monday_t through Wednesday_t, the RMSE for the AR benchmark is 569,418 (642,032) for non-seasonally (seasonally) adjusted UI. Beginning with the nowcast on Thursday_t, the AR model is based on the first (instead of the second) lag of UI, as the UI_{t-1} figure becomes available on Thursday_t. This leads to a substantial reduction in RMSE to 352,507 (445,565) for the non-seasonally (seasonally) adjusted case.

The fourth through eleventh columns of Table 3 report the RMSE ratio for the model in the column heading vis-à-vis the AR benchmark. The ratios are all below one, so that the now- and backcasts based on the machine-learning methods—which incorporate the information in the daily GT terms—always outperform the AR benchmark in terms of RMSE. The improvements in predictive accuracy are typically sizable. For the initial nowcast formed on Monday_t, which is formed ten days before the UI release, the reductions in RMSE relative to the AR benchmark range from approximately 15% to 30% (10% to 20%) for non-seasonally (seasonally) adjusted UI. The RMSE ratios decrease close to monotonically as we move from the nowcast formed on Monday_t to the backcast formed on Monday_{t+1}, which uses

[17] We assess the accuracy of the now- and backcasts using revised UI data.

the largest data overlap (seven days) between the daily GT terms and weekly UI. For the Monday$_{t+1}$ backcast, the reductions in RMSE vis-à-vis the AR benchmark range from around 50% to over 60% (40% to 55%) for non-seasonally (seasonally) adjusted UI. For seasonally adjusted UI, there are often additional improvements in RMSE for the backcasts formed on Tuesday$_{t+1}$ and Wednesday$_{t+1}$. Overall, the predictions generally become more accurate in Table 3 as we include additional days of GT data, with the reduction in MSFE vis-à-vis the AR benchmark reaching as high as 63.2% (for the NN2 backcast formed on Monday$_{t+1}$ for non-seasonally adjusted UI).

The different machine-learning methods in Table 3 perform reasonably similarly, although a distinct pattern emerges with respect to non-seasonally and seasonally adjusted UI. For the latter, as indicated by the bold entries in Table 3, predictions based on linear models nearly always perform the best, suggesting that a linear specification is adequate for capturing the information in the GT terms when it comes to predicting seasonally adjusted UI. In contrast, the most accurate predictions for non-seasonally adjusted UI are almost always based on ANNs, so that accommodating nonlinearities appears important in this case.

The ensembles generally perform well in Table 3. In particular, Ensemble-All—which is an average of the LASSO, ENet, NN1, NN2, and NN3 predictions—often produces close to the lowest RMSE for the individual now- and backcasts. Because we cannot know a priori the best method, the Ensemble-All approach provides a promising practical strategy for now- and backcasting UI.

5.2. COVID-19 Crisis

Figure 2 depicts selected now- and backcasts and realized UI values for the first week of March 2020 through the third week of April 2020, near the advent of the COVID-19 crisis. According to the upper panels, the AR benchmark model has substantial difficulty tracking UI at the start of the crisis. For both the non-seasonally and seasonally adjusted cases, the Monday nowcast (which is based on the second lag of UI) anticipates very little of the

sharp increases in the third and fourth weeks of March. Although the prediction for the second week of April is quite accurate, that for the third week of April massively overstates the actual value. The Monday backcast (which is based on the first lag of UI) anticipates little of the sharp increase in the third week of March and still significantly understates the increase in the next week. The nowcast then significantly overstates UI in the first two weeks of April, while the prediction is much more accurate for the third week of April.[18]

The models that incorporate information from the GT trends are typically much more accurate than the AR benchmark in Figure 2. The middle panels show Monday, Tuesday, and Thursday nowcasts and Monday backcasts for linear models fitted via the LASSO. Apart from the second week of April, the Monday and Tuesday nowcasts perform much better than the corresponding AR benchmark (Monday nowcast) in the upper panels. The inclusion of more timely GT trends as they become available clearly improves the performance of the models, so that the Tuesday nowcasts and Monday backcasts in the middle panels are substantially more accurate than the corresponding AR benchmark (Monday backcast) in the upper panels. The results for the predictions based on the fitted NN2 models in the lower panels follow a similar pattern, although the Monday and Tuesday nowcasts evince larger errors in the first two weeks of April vis-à-vis those in the middle panels.

Compared to the other predictions in Figure 2, the Monday nowcast based on the fitted NN2 model in the lower-left panel is better able to predict the spike in UI in the third week of March. This is because the GT terms play a relatively more important role in the fitted NN2 model for the Monday backcast when it is trained using data before the start of the COVID-19 crisis. The sharp increases in search volume for many GT terms near the start of the crisis thus translate into a comparatively larger UI prediction for the third week of March for the Monday backcast in the lower-left panel. When the training sample includes the UI observation for the third week of March, the GT terms become much more important in the

[18] In the upper panels of Figure 2, the Monday nowcast is the same as the Tuesday and Wednesday nowcasts; the Monday backcast is the same as the Thursday through Saturday nowcasts and Sunday, Tuesday, and Wednesday backcasts.

fitted models, which substantially improves the accuracy of the predictions. This highlights the value of retraining the model each week in our rolling-window estimation procedure. Section 6 analyzes the importance of the GT terms in models fitted using samples that exclude and include data from the crisis.

Next, we investigate whether the relative performance of competing models vis-à-vis the AR benchmark varies systematically with measures of social conditions related to the COVID-19 crisis. Suppose that we want to test whether the difference between MSEs for the AR benchmark and a competing model changes with a measure of social conditions, denoted by z_t. Following Clark and West (2007), we first define the period-t adjusted loss differential for predictive model j:

$$d_t^{(j)} = \left(\text{UI}_t - \widehat{\text{UI}}_t^{\text{AR}}\right)^2 - \left(\text{UI}_t - \widehat{\text{UI}}_t^{(j)}\right)^2 + \left(\widehat{\text{UI}}_t^{\text{AR}} - \widehat{\text{UI}}_t^{(j)}\right)^2 \text{ for } t = 1, \ldots, T, \tag{5.2}$$

where $\widehat{\text{UI}}_t^{\text{AR}}$ and $\widehat{\text{UI}}_t^{(j)}$ are the predictions based on the AR benchmark and model j, respectively. We then use the following regression to test whether the difference between MSEs for the benchmark and model j varies with z_t:

$$d_t^{(j)} = \phi_0^{(j)} + \phi_1^{(j)} z_t + e_t^{(j)}, \tag{5.3}$$

where $e_t^{(j)}$ is a zero-mean error term. If $\phi_1^{(j)}$ equals zero, then the MSE difference does not vary with z_t. Alternatively, if, say, $\phi_1^{(j)} > 0$, then the MSE difference increases as z_t increases (i.e., the AR benchmark becomes less accurate relative to model j as z_t increases).[19]

We use the following four variables for z_t in Equation (5.3):

[19]The first two terms on the right-hand-side of Equation (5.2) constitute the conventional loss differential (Diebold and Mariano 1995; West 1996). However, tests based on the conventional loss differential are subject to severe size distortions when comparing predictions from nested models (Clark and McCracken 2001; McCracken 2007). Clark and West (2007) include the last term on the right-hand-side of Equation (5.2) to improve the size properties of tests based on the loss differential for nested models. In our application, because the competing models include lagged UI, they essentially nest the AR benchmark model, so that we use the adjusted loss differential to help guard against size distortions. Inferences regarding $\phi_1^{(j)}$ in Table 4 are similar when we use the conventional loss differential in Equation (5.3).

- *UI minus pre-crisis average*. Difference between week-t UI and its pre-crisis average, where the pre-crisis average is based on data for the first week of January 2005 through the second week of March 2020.

- *Change in COVID-19 deaths*. Weekly change in reported US deaths due to COVID-19. Data for reported US deaths due to COVID-19 are from the European Centre for Disease Prevention and Control.

- *Stringency index*. Indicator variable that takes a value of one when the government imposes stringent measures to mitigate COVID-19 and zero when such measures are not in place or easing. The indicator variable is based on the stringency index from the Oxford COVID-19 Government Response Tracker (Hale et al. 2020). The index reflects the strength of government actions in response to the COVID-19 pandemic. It increases when measures such as school and retail establishment closures and/or travel restrictions are imposed, and decreases when they are relaxed.

- *Workplace index*. Indicator variable that takes a value of one when the government imposes stringent workplace closure measures to mitigate COVID-19 and zero when such measures are not in place or easing. The indicator variable is based on the workplace closure sub-index of the Hale et al. (2020) stringency index.

Table 4 reports ordinary least squares estimates of $\phi_0^{(j)}$ and $\phi_1^{(j)}$ ($\hat{\phi}_0^{(j)}$ and $\hat{\phi}_1^{(j)}$, respectively) in Equation (5.3) for the Monday backcasts.[20] The second and third columns report results for the UI deviation from its pre-crisis average. In this case, $\phi_0^{(j)}$ can be interpreted as the (adjusted) MSE difference during "normal" times (i.e., before the COVID-19 crisis), while $\phi_1^{(j)}$ measures the change in the MSE difference as UI moves above its pre-crisis average. For all of the competing models and both non-seasonally and seasonally adjusted UI, the $\hat{\phi}_0^{(j)}$ estimate in the second column is insignificant at conventional levels, so that the competing models do not outperform the AR benchmark when UI is at its pre-crisis average. In contrast,

[20]The results are qualitatively similar for the other now- and backcasts.

the $\hat{\phi}_1^{(j)}$ estimates in the third column are all positive and significant (at the 1% level), so that the competing models deliver significant improvements in predictive accuracy vis-à-vis the AR benchmark as UI rises above its pre-crisis level. The results for the change in COVID-19 deaths in the fourth and fifth columns are similar to those in the second and third columns. According to the $\hat{\phi}_0^{(j)}$ estimates in the fourth column, there is no significant difference in MSEs when the change in COVID-19 deaths is zero, while the $\phi_1^{(j)}$ estimates in the fifth column are all positive and significant (at the 1% level). The competing models thus become significantly more accurate relative to the AR benchmark as the change in COVID-19 deaths increases.

The sixth and seventh (eighth and ninth) columns of Table 4 report results for the indicator variable based on the stringency (workplace) index. For both indicator variables, the $\hat{\phi}_0^{(j)}$ estimates in the sixth and eighth columns are all positive and significant (at the 1% level), meaning that the competing models deliver a significantly lower MSE than the AR benchmark when the government does not impose or eases stringent measures to mitigate COVID-19. The $\hat{\phi}_1^{(j)}$ estimates in the seventh and ninth columns are all positive and significant (at the 5% level); hence, the competing models generate further improvements in predictive accuracy vis-à-vis the benchmark when the government imposes stringent measures to combat COVID-19.[21]

Overall, the results in Table 4 indicate that the information in the GT terms significantly improves predictive accuracy during periods when COVID-19 creates greater social stress, as captured by increases in UI relative to its pre-crisis level, increases in COVID-19 deaths, and the imposition of stringent measures to mitigate COVID-19.[22]

[21] Using a χ^2-statistic, the joint hypothesis that $\phi_0^{(j)} = \phi_1^{(j)} = 0$ (i.e., equal predictive ability) is always rejected (at the 1% level) for all choices of z_t.

[22] We also examine whether the inclusion of the following set of ten daily macro-financial variables in the models further improves predictive accuracy: S&P 500 return; Chicago Board Option Exchange volatility index (VIX); gold return; TED spread (difference between the three-month LIBOR and three-month Treasury bill yields); term spread (difference between the ten-year Treasury bond and three-month Treasury bill yields); expected five-year inflation rate (rate at which the five-year Treasury note and five-year TIPS achieve the same yield); default spread (difference between Baa- and Aaa-rated corporate bond yields); Wilshire US real estate investment trust return; Baker, Bloom, and Davis (2016) newspaper-based economic policy uncertainty index; Baker et al. (2020) newspaper-based infectious disease equity market volatility

6. Interpreting the Fitted Models

While the LASSO and ENet facilitate the interpretation of fitted linear models by performing variable selection, fitted ANNs are black boxes that are difficult to interpret. In this section, we use PDPs (Friedman 2001) and variable-importance measures (Greenwell, Boehmke, and McCarthy 2018) to peer into the black box of the fitted ANNs and compare them to fitted linear models.

6.1. Partial-Dependence Plots and Variable Importance

For ease of exposition, we gather the predictors in the 722×1 vector,

$$\boldsymbol{x}_t^{(j)} = \begin{bmatrix} \mathrm{UI}_{t-1(2)} & \boldsymbol{g}_t^{(j)\prime} \end{bmatrix}'. \tag{6.1}$$

Furthermore, we denote the $722 \times \mathcal{T}$ data matrix for the training sample by

$$\boldsymbol{X}_{\mathcal{T}}^{(j)} = \begin{bmatrix} \boldsymbol{x}_1^{(j)} & \cdots & \boldsymbol{x}_{\mathcal{T}}^{(j)} \end{bmatrix}. \tag{6.2}$$

Suppose that we are interested in analyzing the marginal effect of a given predictor, $x_s^{(j)}$, on the expected value of UI_t for a fitted model. Letting $\boldsymbol{x}_{C(s)}^{(j)} = \boldsymbol{x}^{(j)} \setminus x_s^{(j)}$, the partial dependence for $x_s^{(j)}$ is defined as

$$\begin{aligned} \mathrm{PD}\big(x_s^{(j)}\big) &= \mathbb{E}_{\boldsymbol{x}_{C(s)}^{(j)}}\left[\hat{f}^{(j)}\left(x_s^{(j)}, \boldsymbol{x}_{C(s)}^{(j)}\right)\right] \\ &= \int_{\boldsymbol{x}_{C(s)}^{(j)}} \hat{f}^{(j)}\left(x_s^{(j)}, \boldsymbol{x}_{C(s)}^{(j)}\right) p_{C(s)^{(j)}}\left(\boldsymbol{x}_{C(s)}^{(j)}\right) d\boldsymbol{x}_{C(s)^{(j)}}, \end{aligned} \tag{6.3}$$

tracker. For both non-seasonally and seasonally adjusted UI, the inclusion of macro-financial variables offers at best modest improvements in predictive accuracy; in many cases, predictive accuracy deteriorates. When in comes to predicting UI, the daily GT terms thus appear to subsume the relevant information in the daily macro-financial variables.

where

$$p_{C(s)^{(j)}}\left(\boldsymbol{x}_{C(s)}^{(j)}\right) = \int_{x_s^{(j)}} p\left(\boldsymbol{x}^{(j)}\right) dx_s^{(j)}, \tag{6.4}$$

$p\left(\boldsymbol{x}^{(j)}\right)$ is the joint probability density for $\boldsymbol{x}^{(j)}$, and $\hat{f}^{(j)}\left(\boldsymbol{x}^{(j)}\right)$ is the prediction function for the fitted model. Equation (6.3) gives the marginal relationship between the expected value of the target and $x_s^{(j)}$. It is estimated via Monte Carlo integration using the training sample $\boldsymbol{X}_{\mathcal{T}}^{(j)}$:

$$\widehat{\mathrm{PD}}\left(x_s^{(j)}\right) = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \hat{f}^{(j)}\left(x_s^{(j)}, \boldsymbol{x}_{C(s),t}^{(j)}\right), \tag{6.5}$$

where Equation (6.5) is evaluated at the training sample values of $x_s^{(j)}$ (i.e., $x_{s,t}^{(j)}$ for $t = 1, \ldots, \mathcal{T}$) or a set of quantiles.

Of course, the PDP for a fitted linear model will have a constant slope, while it will be a horizontal line for a predictor that is not selected by the LASSO or ENet. By comparing the PDPs for the fitted ANNs to those for the fitted linear models, we can gauge the relative importance of nonlinearities in the former.

Greenwell, Boehmke, and McCarthy (2018) develop a variable-importance metric based on Equation (6.5):

$$\hat{\mathcal{I}}\left(x_s^{(j)}\right) = \left\{ \frac{1}{\mathcal{T}-1} \sum_{t=1}^{\mathcal{T}} \left[\widehat{\mathrm{PD}}\left(x_{s,t}^{(j)}\right) - \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \widehat{\mathrm{PD}}\left(x_{s,t}^{(j)}\right) \right]^2 \right\}^{0.5}. \tag{6.6}$$

Equation (6.6) measures the importance of a predictor via the variation in the PDP around its average value (i.e., the standard deviation). For a predictor with a horizontal PDP, the expected value of the target does not vary with the predictor, so that its variable importance is zero. As the conditional expectation fluctuates more about its average value, the variable importance measure increases. To facilitate comparison across predictors, we scale

Equation (6.6) using the sum of the individual measures:

$$\tilde{\mathcal{I}}\left(x_s^{(j)}\right) = \frac{\hat{\mathcal{I}}\left(x_s^{(j)}\right)}{\sum_{p=1}^{P} \hat{\mathcal{I}}\left(x_p^{(j)}\right)}, \tag{6.7}$$

where P is the total number of predictors, so that $\tilde{\mathcal{I}}\left(x_s^{(j)}\right)$ ranges from zero to one.

6.2. Importance of Google Trends Terms

Figures 3 and 4 depict variable-importance measures based on Equation (6.7) for the top 25 predictors for linear models fitted via the LASSO and fitted NN2 models, respectively, for non-seasonally adjusted UI. The left panels of the figures correspond to models fitted with data through the end of 2019, before the start of the COVID-19 crisis; the right panels are for models estimated using data through the last week of July 2020, so that the training sample includes the crisis. Note that a given GT term can appear up to seven times in the same plot, due to our mixed-frequency framework. To conserve space, we focus on fitted models for the Monday, Tuesday, and Thursday nowcasts and Monday backcast (with data overlaps of zero, one, three, and seven days, respectively).

For the pre-crisis sample, the lag of UI (*lag*) is the most important predictor for both the LASSO and NN2 in Figures 3 and 4. For the linear models fitted via the LASSO in Figure 3, the $\tilde{\mathcal{I}}\left(x_s^{(j)}\right)$ scores for *lag* range from approximately 0.25 to 0.45. Taken together, the GT terms are more important in the fitted NN2 models in Figure 4. Nevertheless, *lag* still predominates, with $\tilde{\mathcal{I}}\left(x_s^{(j)}\right)$ scores ranging from around 0.06 to 0.12.

The GT terms become substantively more important in the right panels of Figures 3 and 4 when we include data from the COVID-19 crisis. Although *lag* remains the most or next-to-most important predictor for the Monday, Tuesday, and Thursday nowcasts across both models, it drops out of the top 25 for the Monday backcast. In other words, when we use the maximum data overlap in computing the predictions, the fitted models assign little importance to the autocorrelation in UI, so that the GT terms dominate the AR component.

Table 5 provides additional information on the growing importance of the GT terms in the sequence of now- and backcasts, especially for the sample that includes the COVID-19 crisis. The table reports the joint importance of the GT terms for each prediction-formation day (for non-seasonally adjusted UI). For the linear models fitted via the LASSO and pre-crisis sample in the third column, the GT terms grow in importance from around 0.60 for the initial nowcasts to 0.75 for the later backcasts. When the training sample includes the COVID-19 crisis (see the fourth column), the GT terms again grow in importance, but the level is markedly higher for each prediction-formation day. For the Monday nowcast, the joint importance score for the GT terms is 0.715. It reaches 1.000 for the Saturday nowcast through the Wednesday backcast, so that the AR component becomes unimportant. A similar pattern holds for the fitted NN2 models in the last two columns, with the joint importance measures for the GT terms nearly always larger (often substantially so) than the corresponding values in the third and fourth columns.

The increasing importance of the GT terms since the start of the COVID-19 crisis is also evident in Figure 5, which shows the number of predictors selected by the LASSO and ENet for rolling-window estimation of the linear predictive models underlying the Monday, Tuesday, and Thursday nowcasts and Monday backcast. In general, the number of selected predictors sharply increases in the fitted linear models with the advent of the crisis, with spikes evident in the number of selected predictors for training samples ending in the third week of March 2020. As expected, the ENet usually selects more predictors than the LASSO in Figure 5.

Returning to Figures 3 and 4, for the prediction-formation days where the GT terms matter the most—most notably, the Monday backcast—there is a tendency for both the fitted linear and NN2 models to place more weight on recently available search queries. For example, the Monday backcast attaches the greatest importance to GT terms for Saturday and Friday (recall the two-day lag in the availability of the GT data). To explore this issue further, Figure 6 shows heatmaps for the joint importance of the GT terms organized

according to the day of the week. For the pre-crisis sample in the left panels, there is no discernible pattern in the importance of the GT terms across the days of the week. For the sample that includes the COVID-19 crisis in the right panels, a strong pattern is evident: with the exception of the Wednesday backcast, the now- and backcasts attach relatively high importance to GT terms for the most recent day of available data. This is most evident for the Monday backcast, which uses the maximum data overlap of seven days, where the collective importance of GT terms for Saturday is 0.73 (0.46) for the linear model fitted via the LASSO (fitted NN2 model). This highlights the usefulness of mixed-frequency GT data for anticipating UI during the COVID-19 crisis.

Looking back to Figures 3 and 4, an interesting pattern emerges in the types of GT terms that appear important across the two samples. For the fitted linear and NN2 models, the pre-COVID-19 sample is characterized by a wide variety of search queries (e.g., some geographical terms and some generic terms like *workers compensation*). For the sample that includes the COVID-19 crisis, there appears to be greater emphasis on search queries related to the application process for unemployment insurance benefits (e.g., *how to file for unemployment*, *unemployment application*, and *unemployment office*).

6.3. Nonlinearities

To get a sense of the strength of the nonlinearities in the fitted ANNs, we investigate PDPs for some of the most relevant predictors. We again report results for linear models fitted via the LASSO and fitted NN2 models for training samples that exclude and include the COVID-19 crisis. Figure 7 presents PDPs for the AR component (*lag*), *how to file for unemployment*, and *unemployment office illinois*. The figure reports results for the Thursday nowcast and Monday backcast (for non-seasonally adjusted UI). The two GT terms are for the most recently available day of GT data. The first GT term, *how to file for unemployment*, is included because it is the most important predictor in the fitted linear and NN2 models for the Monday backcast during the sample that includes the crisis. We include *unemployment*

office illinois because it provides an example of an important geographical search query that enters the top 25 for the Monday backcast.[23] Note that Figure 7 uses different scales for the different training samples, as average UI is much higher for the sample that includes the COVID-19 crisis.

By construction, the PDPs are linear for the linear models fitted via the LASSO. For the Thursday nowcast, the predictive relationships for the AR component (*lag*) are quite close to linear for the fitted NN2 models. In other words, although the NN2 allows for nonlinear relationships, when we train the model using available data, the relationship involving the AR component is essentially linear. For the Monday backcast, there is more evidence of a nonlinear relationship for the NN2 model fitted using the sample that includes the COVID-19 crisis, with the slope becoming flatter for higher values of *lag*.

With respect to *how to file for unemployment* and *unemployment office illinois* in Figure 7, nonlinearities are evident for the fitted NN2 models when the training sample includes data from the COVID-19 crisis, especially for the Monday nowcast. Specifically, for the fitted NN2 models based on data that include the crisis in the lower panels, the slopes of the PDP curves become notably steeper for large values of the GT terms, so that the response of non-seasonally adjusted UI to the GT terms becomes stronger for more extreme values for the latter.[24]

7. Conclusion

We show that the information in high-dimensional daily internet search-volume data can be used to substantially improve predictions of weekly UI in anticipation of its Thursday release by the Department of Labor. We construct a sequence of now- and backcasts that are formed ten days to one day ahead of the UI release on Thursday of each week. To effectively

[23] At the end of August 2020, Illinois had the sixth (seventh) highest number of confirmed cases (deaths) across US states (https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html).

[24] The PDPs for the fitted ANNs for seasonally adjusted UI show little evidence of nonlinearities, so that a linear specification appears adequate for modeling the predictive relationships between the GT terms and seasonally adjusted UI.

utilize the information in a large number of daily GT terms related to *unemployment*, we estimate the predictive models underpinning the now- and backcasts using machine-learning techniques in a mixed-frequency framework. The mixed-frequency framework allows us to incorporate daily GT data as they become available, thereby providing us with more timely information for predicting weekly UI, while machine-learning methods are appropriate for our high-dimensional setting. Our mixed-frequency approach also allows us to measure the predictive power of each additional day of information. In a simulated out-of-sample exercise, now- and backcasts based on daily GT terms substantially outperform an AR benchmark in terms of RMSE. As the sequence of now- and backcasts incorporates more recent daily GT data, predictive accuracy generally improves, leading to reductions in RMSE of up to 63% vis-à-vis the AR benchmark.

We detect strong links between the daily GT terms and COVID-19 crisis. The daily GT terms are especially useful for improving the accuracy of the now- and backcasts near the advent of the crisis. We also find that the predictive accuracy of models that include the GT terms improves significantly relative to that of the AR benchmark model as social conditions associated with the crisis worsen. Variable-importance measures for the fitted models reveal that the GT terms become more relevant when the training samples include data from the crisis. Variable-importance measures also show that GT terms for the most recently available day are typically the most germane in the fitted models, highlighting the value of our mixed-frequency approach.

We are in the process of creating a website that will provide updated, real-time now- and backcasts of UI on a daily basis using the methods developed in this paper.[25] We are also in the process of using our methodology to generate now- and backcasts of weekly UI for individual US states, as well as predictions of similar measures of unemployment benefit claims for a number of European countries.

[25] Available at https://www.uinowcast.org/.

References

Aaronson, D., S. A. Brave, R. A. Butters, D. W. Sacks, and B. Seo (2020). Using the Eye of the Storm to Predict the Wave of Covid-19 UI Claims. Federal Reserve Bank of Chicago Working Paper No. 2020-10.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring Economic Policy Uncertainty. *Quarterly Journal of Economics* 131:4, 1593–1636.

Baker, S. R., N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin (2020). The Unprecedented Stock Market Reaction to COVID-19. *Review of Asset Pricing Studies* 10:9, 742–758.

Barron, A. R. (1994). Approximation and Estimation Bounds for Artificial Neural Networks. *Machine Learning* 14:1, 115–133.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009). Simultaneous Analysis of Lasso and Dantzig Selector. *Annals of Statistics* 37:4, 1705–1732.

Borup, D. and E. C. M. Schütte (forthcoming). In Search of a Job: Forecasting Employment Growth Using Google Trends. *Journal of Business and Economic Statistics*.

Brave, S. A., R. A. Butters, and A. Justiniano (2019). Forecasting Economic Activity with Mixed Frequency BVARs. *International Journal of Forecasting* 35:4, 1692–1707.

Choi, H. and H. Varian (2012). Predicting the Present with Google Trends. *Economic Record* 88:S1, 2–9.

Clark, T. E. and M. W. McCracken (2001). Tests of Equal Forecast Accuracy and Encompassing for Nested Models. *Journal of Econometrics* 105:1, 85–110.

Clark, T. E. and K. D. West (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics* 138:1, 291–311.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States. *Journal of Business and Economic Statistics* 26:4, 546–554.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics* 6:1, 3–73.

Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems* 2:4, 303–314.

D'Amuri, F. and J. Marcucci (2017). The Predictive Power of Google Searches in Forecasting US Unemployment. *International Journal of Forecasting* 33:4, 801–816.

Davidson, P. (2020). Unemployment Claims Figures Could Be Much Lower Because of New Seasonal Adjustment Approach. *USA Today*, September 3.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13:3, 253–263.

Diebold, F. X. and M. Shin (2019). Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian Lasso and its Derivatives. *International Journal of Forecasting* 35:4, 1679–1691.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 178:1, 57–82.

Foroni, C. and M. Marcellino (2014). A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates. *International Journal of Forecasting* 30:3, 554–568.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* 29:5, 1189–1232.

Funahashi, K.-I. (1989). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks* 2:3, 183–192.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a Risk-Return Trade-Off After All. *Journal of Financial Economics* 76:3, 509–548.

Goldsmith-Pinkham, P. and A. Sojourner (2020). Predicting Initial Unemployment Insurance Claims Using Google Trends. Working Paper.

Greenwell, B. M., B. C. Boehmke, and A. J. McCarthy (2018). A Simple and Effective Model-Based Variable Importance Measure. Working Paper (arXiv:1805.04755v1).

Hale, T., N. Angrist, E. Cameron-Blake, L. Hallas, B. Kira, S. Majumdar, A. Petherick, T. Phillips, H. Tatlow, and S. Webster (2020). Oxford COVID-19 Government Response Tracker. Blavatnik School of Government.

Hastie, T. and J. Qian (2016). Glmnet Vignette. Working Paper.

Hoerl, A. E. and R. W. Kennard (1970). Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* 12:1, 69–82.

Hornik, K. (1991). Approximation Capabilities of Multilayer Feedforward Networks. *Neural Networks* 4:2, 251–257.

Hornik, K., M. Stinchcombe, and H. White (1989). Multilayer Feedforward Networks Are Universal Approximators. *Neural Networks* 2:5, 359–366.

Hui, F. K. C., D. I. Warton, and S. D. Foster (2015). Tuning Parameter Selection for the Adaptive Lasso Using ERIC. *Journal of the American Statistical Society* 110:509, 262–269.

Kingma, D. P. and J. Ba (2015). *Adam: A Method for Stochastic Optimization*. Third Annual International Conference on Learning Representations. San Diego.

Kotchoni, R., M. Leroux, and D. Stevanovic (2019). Macroeconomic Forecast Accuracy in a Data-Rich Environment. *Journal of Applied Econometrics* 34:7, 1050–1072.

Larson, W. D. and T. M. Sinclair (2020). Nowcasting Unemployment Insurance Claims in the Time of COVID-19. CAMA Working Paper No. 63/2020.

Lewis, D., K. Mertens, and J. H. Stock (2020). US Economic Activity During the Early Weeks of the SARS-Cov-2 Outbreak. NBER Working Paper No. 26954.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. Boston, MA: Academic Press.

McCracken, M. W. (2007). Asymptotics for Out of Sample Tests of Granger Causality. *Journal of Econometrics* 140:2, 719–752.

Medeiros, M. C., G. F. R. Vasconcelos, A. Veiga, and E. Zilberman (forthcoming). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business and Economic Statistics*.

Meinshausen, N. and B. Yu (2009). Lasso-Type Recovery of Sparse Representations for High-Dimensional Data. *Annals of Statistics* 37:1, 246–270.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:3, 703–708.

Niesert, R. F., J. A. Oorschot, C. P. Veldhuisen, K. Brons, and R.-J. Lange (2020). Can Google Search Data Help Predict Macroeconomic Series? *International Journal of Forecasting* 36:3, 1163–1172.

Rinz, K. (2020). Understanding Unemployment Insurance Claims and Other Labor Market Data During the COVID-19 Pandemic. Working Paper.

Rolnick, D. and M. Tegmark (2018). *The Power of Deeper Networks for Expressing Natural Functions*. Sixth Annual International Conference on Learning Representations. Vancouver.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* 6:2, 461–464.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)* 58:1, 267–288.

Timmermann, A. (2006). Forecast Combinations. In *Handbook of Economic Forecasting*. Ed. by G. Elliott, C. W. J. Granger, and A. Timmermann. Vol. 1. Amsterdam: Elsevier, pp. 135–196.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica* 64:5, 1067–1084.

Zhang, C.-H. and J. Huang (2008). The Sparsity and Bias of the Lasso Selection in High-Dimensional Linear Regression. *Annals of Statistics* 36:4, 1567–1594.

Zou, H. and T. Hastie (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67:2, 301–320.

Table 1: GT terms

unemployment	*unemployment benefits nj*	*to file for unemployment*
unemployment benefits	*unemployment insurance benefits*	*how to file unemployment*
unemployment office	*unemployment benefits california*	*file unemployment claim*
unemployment insurance	*unemployment benefits new york*	*file for unemployment benefits*
file for unemployment	*unemployment benefits florida*	*how apply for unemployment*
apply for unemployment	*unemployment florida*	*to apply for unemployment*
unemployment claim	*nys unemployment*	*how to apply for unemployment*
how to file for unemployment	*florida unemployment benefits*	*apply for unemployment benefits*
unemployment application	*office of unemployment*	*texas unemployment*
unemployment compensation	*the unemployment office*	*unemployment in florida*
unemployment number	*state unemployment office*	*unemployment file claim*
unemployment online	*unemployment ky*	*file unemployment*
unemployment rate	*ky unemployment office*	*wisconsin unemployment*
pa unemployment	*unemployment ca*	*oregon unemployment*
claim unemployment	*ca unemployment office*	*edd online*
unemployment nys	*employment office*	*edd*
unemployment ny	*employment*	*my ui*
ohio unemployment	*unemployment office nj*	*application for unemployment*
unemployment texas	*unemployment nj*	*unemployment iowa*
florida unemployment	*unemployment pa*	*wi unemployment*
nj unemployment	*unemployment il*	*pa unemployment compensation*
unemployment oregon	*unemployment office pa*	*workers compensation*
california unemployment	*unemployment office illinois*	*florida unemployment compensation*
unemployment california	*unemployment illinois*	*unemployment compensation benefits*
new york unemployment	*state unemployment insurance*	*pennsylvania unemployment compensation*
unemployment washington	*california unemployment insurance*	*ohio unemployment compensation*
unemployment wisconsin	*unemployment insurance ny*	*pennsylvania unemployment*
unemployment indiana	*ny unemployment*	*unemployment ohio*
unemployment nc	*unemployment health insurance*	*federal unemployment*
ca unemployment	*unemployment insurance nys*	*unemployment alabama*
texas unemployment benefits	*unemployment insurance new york*	*unemployment phone number*
texas benefits	*unemployment insurance tax*	*number for unemployment*
ny unemployment benefits	*unemployment insurance claim*	*filing unemployment online*
unemployment benefits texas	*unemployment insurance nj*	
claim unemployment benefits	*unemployment insurance office*	

The table lists the 103 Google Trends (GT) terms used to construct now- and backcasts of weekly unemployment insurance initial claims.

Table 2: Information flow

(1) Prediction formation	(2) Backcast/ nowcast	(3) Google Trends used for prediction	(4) Data overlap (days)	(5) j	(6) Latest available UI release
Monday_t	Nowcast	Sunday_{t-1} to Saturday_{t-1}	0	7	Week $t-2$
Tuesday_t	Nowcast	Monday_{t-1} to Sunday_t	1	6	Week $t-2$
Wednesday_t	Nowcast	Tuesday_{t-1} to Monday_t	2	5	Week $t-2$
Thursday_t	Nowcast	Wednesday_{t-1} to Tuesday_t	3	4	Week $t-1$
Friday_t	Nowcast	Thursday_{t-1} to Wednesday_t	4	3	Week $t-1$
Saturday_t	Nowcast	Friday_{t-1} to Thursday_t	5	2	Week $t-1$
Sunday_{t+1}	Backcast	Saturday_{t-1} to Friday_t	6	1	Week $t-1$
Monday_{t+1}	Backcast	Sunday_t to Saturday_t	7	0	Week $t-1$
Tuesday_{t+1}	Backcast	Monday_t to Sunday_{t+1}	6	-1	Week $t-1$
Wednesday_{t+1}	Backcast	Tuesday_t to Monday_{t+1}	5	-2	Week $t-1$

The table reports the information flow for daily Google Trends data and data releases of unemployment insurance initial claims (UI) used for now- and backcasts of week-t UI. The first column provides the prediction-formation day, where the subscript denotes the week when the prediction is made. The second column provides the classification as a nowcast or backcast of week-t UI. The third column provides the range of daily Google Trends terms used for prediction. The fourth column provides the number of days of overlap between the Google Trends terms in the third column and week-t UI. The fifth column provides the value for j in Equation (4.1) The sixth column provides the latest release of UI available at the time of prediction formation.

Table 3: RMSE ratios

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Linear		ANN			Ensemble		
Prediction formation	Nowcast/ backcast	AR RMSE	LASSO	ENet	NN1	NN2	NN3	Linear	ANN	All
Panel A: Non-seasonally adjusted UI										
Monday_t	Nowcast	569,418	0.745	0.755	**0.697**	0.841	0.763	0.749	0.758	0.747
Tuesday_t	Nowcast	569,418	0.682	0.684	**0.629**	0.683	0.729	0.681	0.657	0.660
Wednesday_t	Nowcast	569,418	0.626	0.619	0.658	0.628	**0.583**	0.620	0.602	0.603
Thursday_t	Nowcast	352,507	0.648	0.642	0.705	**0.523**	0.627	0.644	0.547	0.588
Friday_t	Nowcast	352,507	0.601	**0.573**	0.677	0.607	0.597	0.587	0.577	0.577
Saturday_t	Nowcast	352,507	0.552	0.543	0.677	**0.433**	0.678	0.546	0.475	0.504
Sunday_{t+1}	Backcast	352,507	0.580	0.587	0.436	0.475	0.450	0.583	**0.421**	0.492
Monday_{t+1}	Backcast	352,507	0.497	0.481	0.383	**0.368**	0.554	0.487	0.389	0.434
Tuesday_{t+1}	Backcast	352,507	0.496	0.491	**0.393**	0.463	0.619	0.493	0.442	0.463
Wednesday_{t+1}	Backcast	352,507	0.533	0.540	**0.447**	0.554	0.607	0.536	0.482	0.501
Panel B: Seasonally adjusted UI										
Monday_t	Nowcast	642,032	0.829	**0.811**	0.974	0.912	0.813	0.820	0.886	0.842
Tuesday_t	Nowcast	642,032	0.715	0.718	0.866	0.923	0.748	**0.714**	0.819	0.757
Wednesday_t	Nowcast	642,032	0.677	**0.670**	0.748	0.742	0.834	0.672	0.734	0.690
Thursday_t	Nowcast	445,565	0.675	**0.660**	0.975	0.689	0.761	0.665	0.774	0.712
Friday_t	Nowcast	445,565	**0.603**	0.617	0.758	0.630	0.643	0.606	0.638	0.618
Saturday_t	Nowcast	445,565	**0.492**	0.513	0.600	0.576	0.720	0.501	0.589	0.532
Sunday_{t+1}	Backcast	445,565	**0.465**	0.527	0.549	0.559	0.597	0.491	0.526	0.504
Monday_{t+1}	Backcast	445,565	**0.454**	0.517	0.535	0.537	0.605	0.483	0.528	0.502
Tuesday_{t+1}	Backcast	445,565	0.464	0.474	0.497	0.534	0.602	**0.461**	0.517	0.485
Wednesday_{t+1}	Backcast	445,565	0.460	0.495	**0.441**	0.538	0.686	0.474	0.496	0.483

The table reports out-of-sample results for now- and backcasts of weekly unemployment insurance initial claims (UI) formed on the day indicated in the first column, where the subscript denotes the week when the prediction is made. The third column reports the root mean squared error (RMSE) for an autoregressive (AR) benchmark model. The fourth through eleventh columns report the RMSE ratio for the model in the column heading vis-à-vis the AR benchmark. The competing models incorporate the information in daily search volumes for Google Trends terms related to *unemployment*. The fourth and fifth columns are for linear models fitted via the LASSO and elastic net (ENet), respectively. The sixth through eighth columns are for fitted artificial neural networks (ANNs) with one (NN1), two (NN2), and three (NN3) hidden layers, respectively. The ensemble in the ninth (tenth) column is an average of the predictions in the fourth and fifth (sixth through eighth) columns. The ensemble in the eleventh column is an average of the predictions in the fourth through eighth columns. The out-of-sample period begins in the first week of January 2015 and ends in the first week of August 2020. Bold indicates the prediction with the lowest RMSE.

Table 4: Relative performance and the COVID-19 crisis

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	UI minus pre-crisis average		Change in COVID-19 deaths		Stringency index		Workplace index	
Model	$\hat{\phi}_0^{(j)}$	$\hat{\phi}_1^{(j)}$	$\hat{\phi}_0^{(j)}$	$\hat{\phi}_1^{(j)}$	$\hat{\phi}_0^{(j)}$	$\hat{\phi}_1^{(j)}$	$\hat{\phi}_0^{(j)}$	$\hat{\phi}_1^{(j)}$
Panel A: Non-seasonally adjusted UI								
LASSO	5.46	18.16***	14.36	0.13***	0.07***	246.60**	0.07***	398.29**
ENet	5.75	17.78***	14.37	0.13***	0.08***	246.84**	0.09***	398.61**
NN1	5.67	16.26***	14.06	0.10***	0.16***	230.31**	0.16***	371.94**
NN2	6.45	17.79***	15.29	0.12***	0.15***	266.67**	0.15***	412.93**
NN3	7.18	21.00***	17.15	0.16***	0.12***	296.55**	0.15***	479.34**
Ensemble-Linear	5.61	17.97***	14.37	0.13***	0.07***	246.72**	0.08***	398.45**
Ensemble-ANN	6.43	18.35***	15.50	0.13***	0.14***	260.84**	0.14***	421.40**
Ensemble-All	6.02	18.16***	14.93	0.13***	0.11***	253.78**	0.11***	409.93**
Panel B: Seasonally adjusted UI								
LASSO	5.46	18.16***	14.36	0.13***	0.07***	246.60**	0.07***	398.29**
ENet	5.75	17.78***	14.37	0.13***	0.08***	246.84**	0.09***	398.61**
NN1	5.67	16.26***	14.06	0.10***	0.16***	230.31**	0.16***	371.94**
NN2	6.45	17.79***	15.29	0.12***	0.15***	266.67**	0.15***	412.93**
NN3	7.18	21.00***	17.15	0.16***	0.12***	296.55**	0.15***	479.34**
Ensemble-Linear	5.61	17.97***	14.37	0.13***	0.07***	246.72**	0.08***	398.45**
Ensemble-ANN	6.43	18.35***	15.50	0.13***	0.14***	260.84**	0.14***	421.40**
Ensemble-All	6.02	18.16***	14.93	0.13***	0.11***	253.78**	0.11***	409.93**

The table reports ordinary least squares parameter estimates for the regression model, $d_t^{(j)} = \phi_0^{(j)} + \phi_1^{(j)} z_t + e_t^{(j)}$, pertaining to the Monday backcast for different z_t variables, where $d_t^{(j)}$ is the adjusted loss differential. The adjusted loss differential corresponds to the week-t loss for a prediction of unemployment insurance initial claims (UI) based on the model in the first column relative to that for an autoregressive benchmark model. The competing models incorporate the information in daily search volumes for Google Trends terms related to *unemployment*. LASSO and ENet are linear models fitted via the LASSO and elastic net (ENet), respectively. NN1, NN2, and NN3 are fitted artificial neural networks (ANNs) with one (NN1), two (NN2), and three (NN3) hidden layers, respectively. Ensemble-Linear (Ensemble-ANN) is an average of the predictions based on the LASSO and ENet (NN1, NN2, and NN3) models; Ensemble-All is the average of the predictions based on the LASSO, ENet, NN1, NN2, and NN3 models. The z_t variable for the regression is given at the top of the table. UI minus pre-crisis average is week-t UI minus its average value for the first week of January 2005 through the second week of March 2020. Change in COVID-19 deaths is based on data from the European Centre for Disease Prevention and Control. Stringency (workplace) index is an indicator variable that takes a value of one when the government imposes stringent measures to mitigate COVID-19 according to the Oxford COVID-19 Government Response Tracker stringency index (workplace closure sub-index) and zero when such measures are not in place or easing. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively, based on t-statistics computed using heteroskedasticity- and autocorrelation-robust standard errors (Newey and West 1987).

Table 5: Importance of GT terms

(1)	(2)	(3)	(4)	(5)	(6)
Prediction formation	Data overlap (days)	LASSO excluding COVID-19 crisis	LASSO including COVID-19 crisis	NN2 excluding COVID-19 crisis	NN2 including COVID-19 crisis
Monday_t	0	0.570	0.715	0.920	0.941
Tuesday_t	1	0.652	0.740	0.892	0.948
Wednesday_t	2	0.654	0.761	0.891	0.834
Thursday_t	3	0.670	0.753	0.895	0.973
Friday_t	4	0.577	0.771	0.892	0.844
Saturday_t	5	0.542	1.000	0.865	0.991
Sunday_{t+1}	6	0.569	1.000	0.862	0.989
Monday_{t+1}	7	0.774	1.000	0.949	0.992
Tuesday_{t+1}	6	0.789	1.000	0.966	1.000
Wednesday_{t+1}	5	0.730	1.000	0.966	0.994

The table reports joint variable-importance measures for all of the daily Google Trends (GT) terms in fitted linear models estimated via the LASSO and fitted artificial neural networks with two hidden layers (NN2). The second column provides the number of days of overlap between the daily Google Trends terms and week-t unemployment insurance initial claims. The third and fifth (fourth and sixth) columns report results for training samples excluding (including) the COVID-19 crisis. Results pertain to non-seasonally adjusted unemployment insurance initial claims.

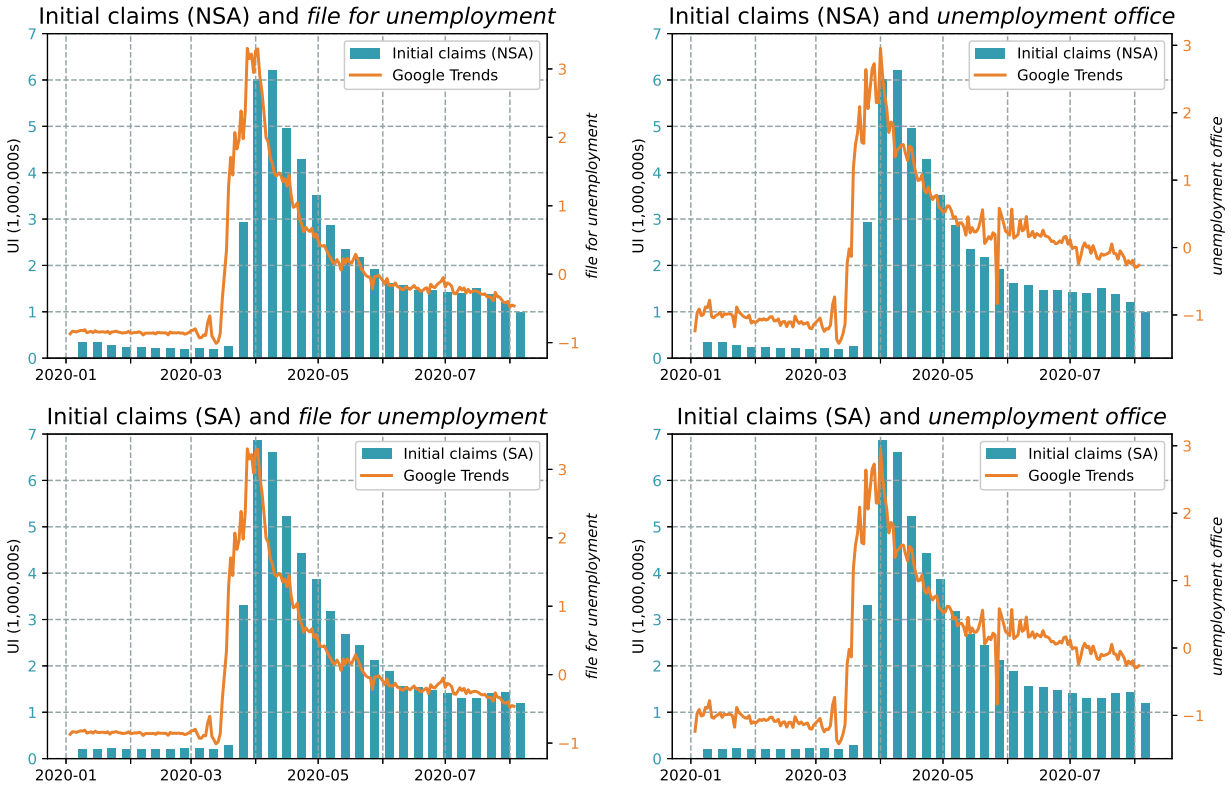


Figure 1: UI and GT terms

The figure depicts non-seasonally adjusted (NSA, upper panels) and seasonally adjusted (SA, lower panels) weekly unemployment insurance initial claims (UI, left axis) at their release date and (standardized) daily search volume for two Google Trends (GT) terms (right axis): *file for unemployment* (left panels) and *unemployment office* (right panels). The sample period spans the first week of January 2020 through the first week of August 2020.

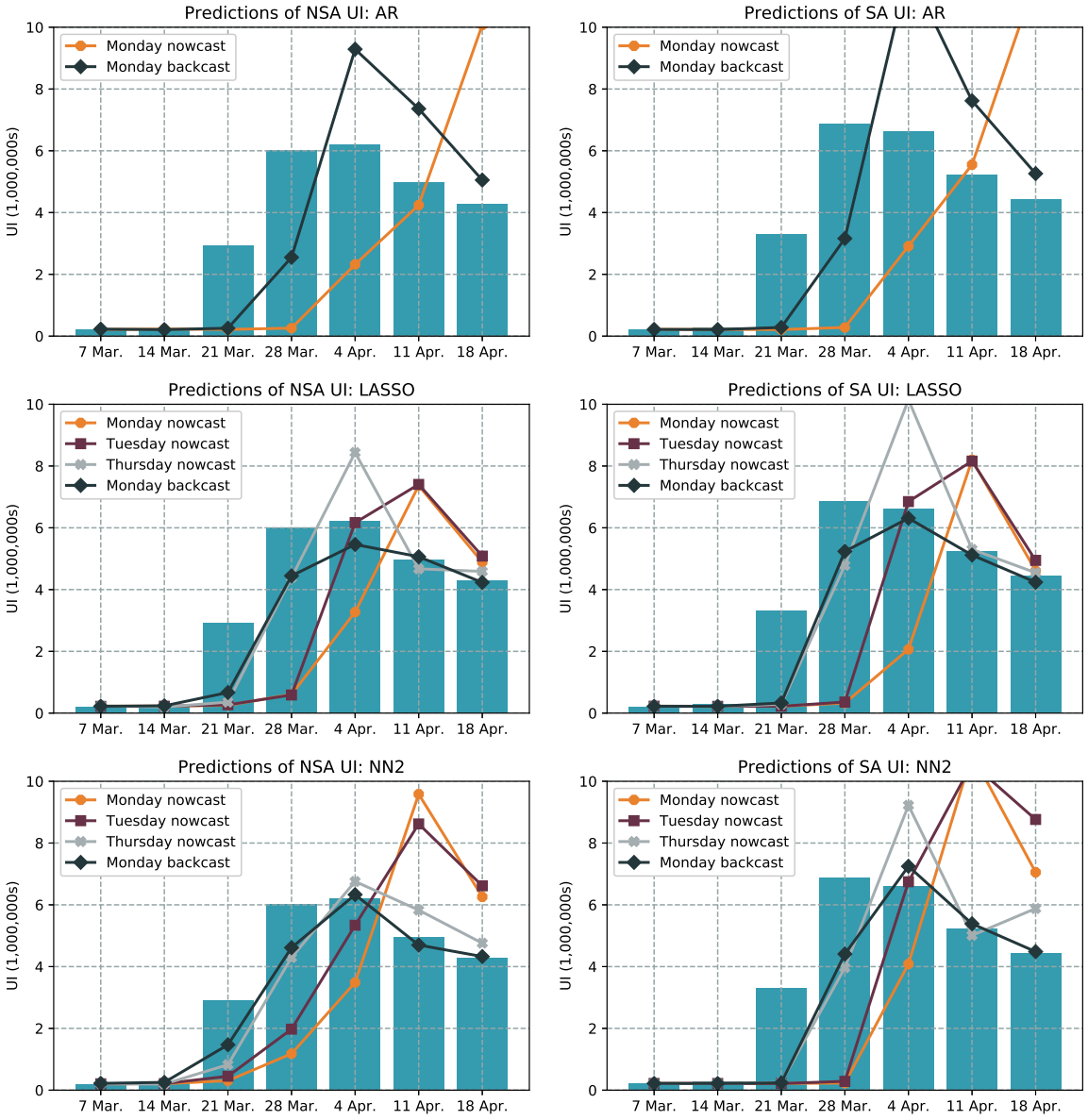


Figure 2: Predictions at the advent of COVID-19

The figure depicts predictions of weekly unemployment insurance initial claims (UI) based on fitted autoregressive (AR) benchmark models (upper panels), linear models fitted via the LASSO (middle panels), and fitted two-layer artificial neural networks (NN2, lower panels). The models in the middle and lower panels incorporate the information in daily search volumes for Google Trends terms related to *unemployment*. Each panel contains results for four prediction-formation days: Monday nowcast, Tuesday nowcast, Thursday nowcast, and Monday backcast. The blue bars show the realized values of UI. Results are reported for non-seasonally adjusted (NSA, left panels) and seasonally adjusted (SA, right panels) UI.

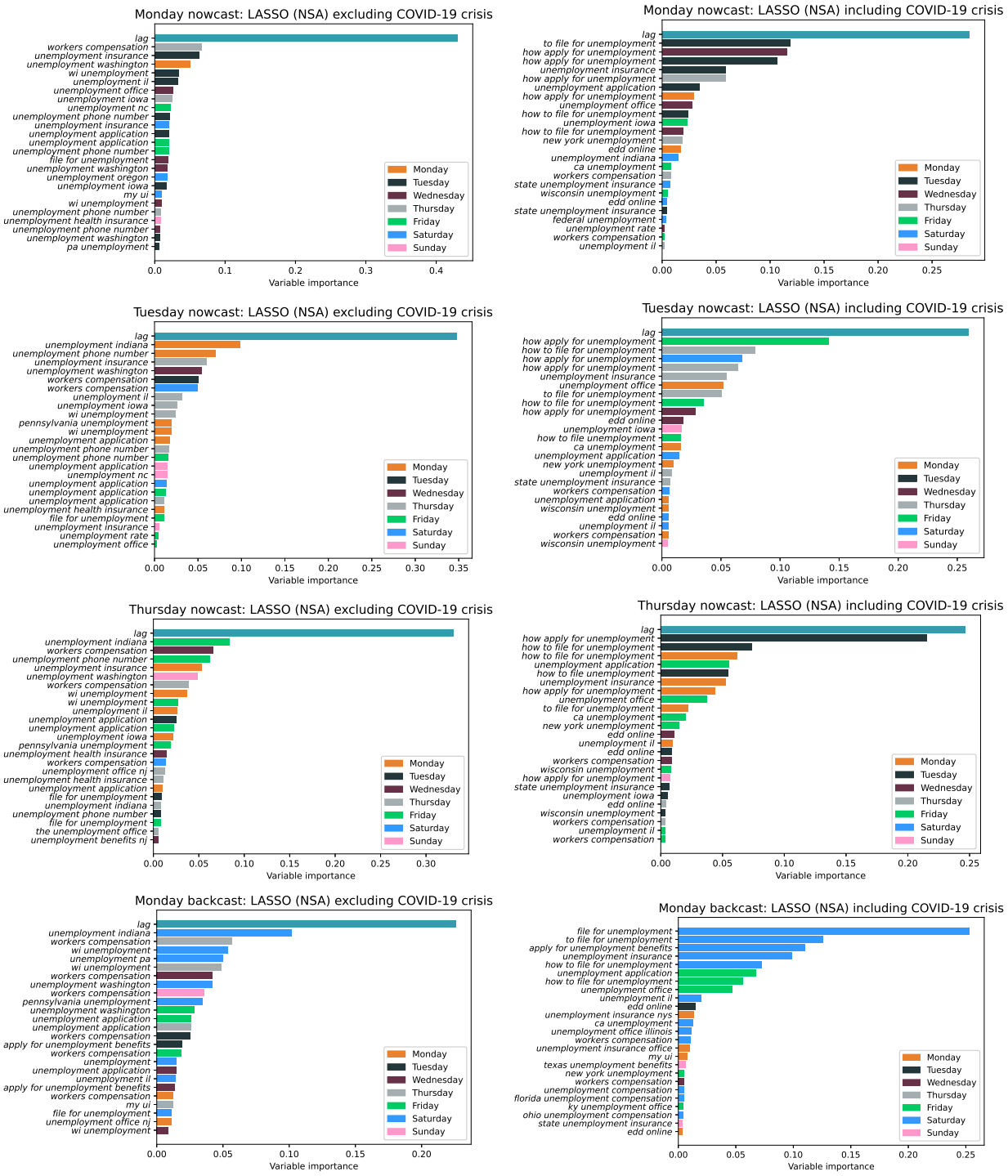


Figure 3: Variable-importance measures for fitted linear models

The figure depicts variable-importance measures for the top 25 predictors for linear models fitted via the LASSO. Results are reported for Monday, Tuesday, and Thursday nowcasts and the Monday backcast. The training sample for the left (right) panels ends in the last week of December 2019 (first week of August 2020), thereby excluding (including) the COVID-19 crisis. Results are reported for non-seasonally adjusted (NSA) weekly unemployment insurance initial claims.

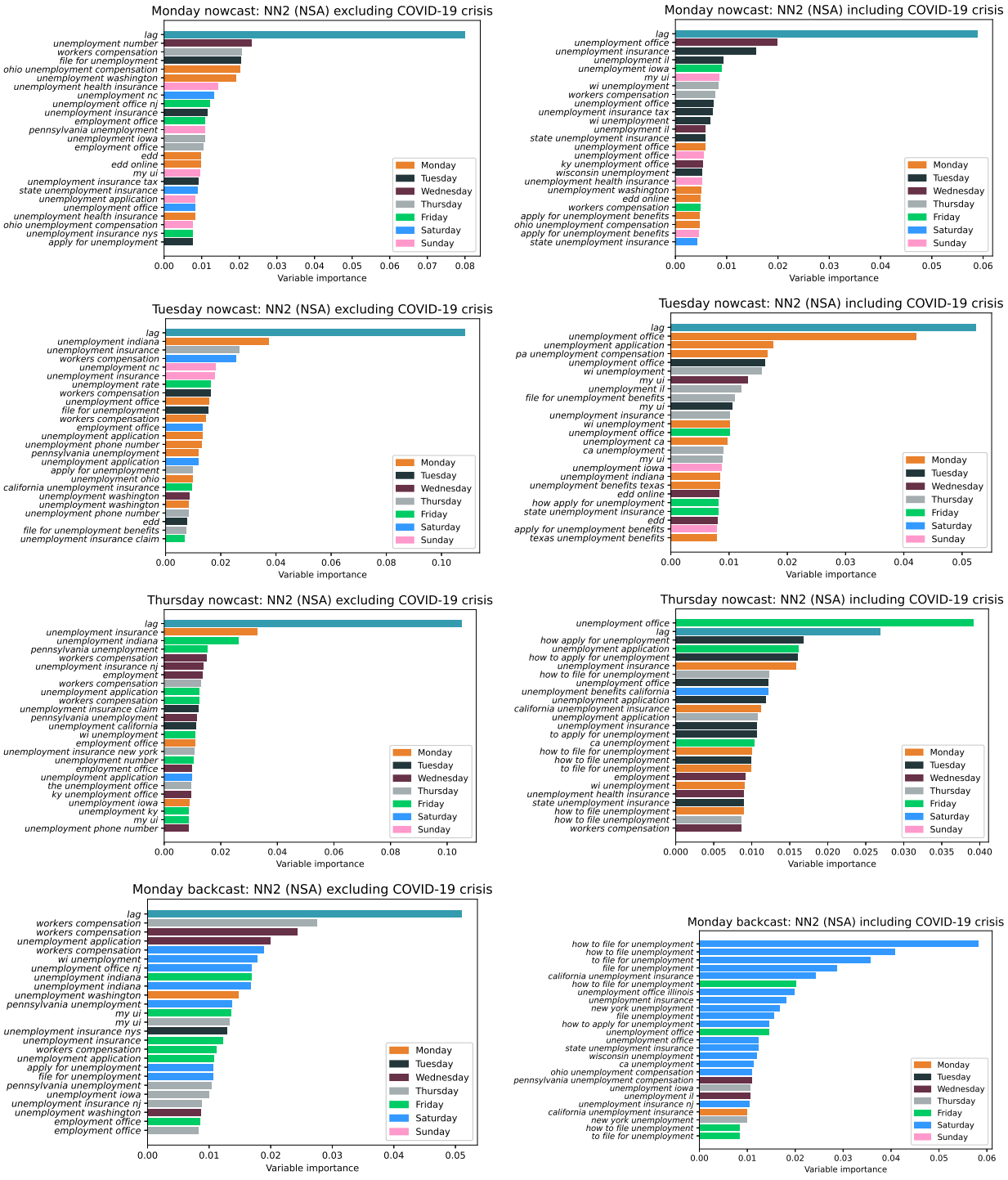


Figure 4: Variable-importance measures for fitted artificial neural networks

The figure depicts variable-importance measures for the top 25 predictors for fitted artificial neural networks with two hidden layers (NN2). Results are reported for Monday, Tuesday, and Thursday nowcasts and the Monday backcast. The training sample for the left (right) panels ends in the last week of December 2019 (first week of August 2020), thereby excluding (including) the COVID-19 crisis. Results are reported for non-seasonally adjusted (NSA) weekly unemployment insurance initial claims.

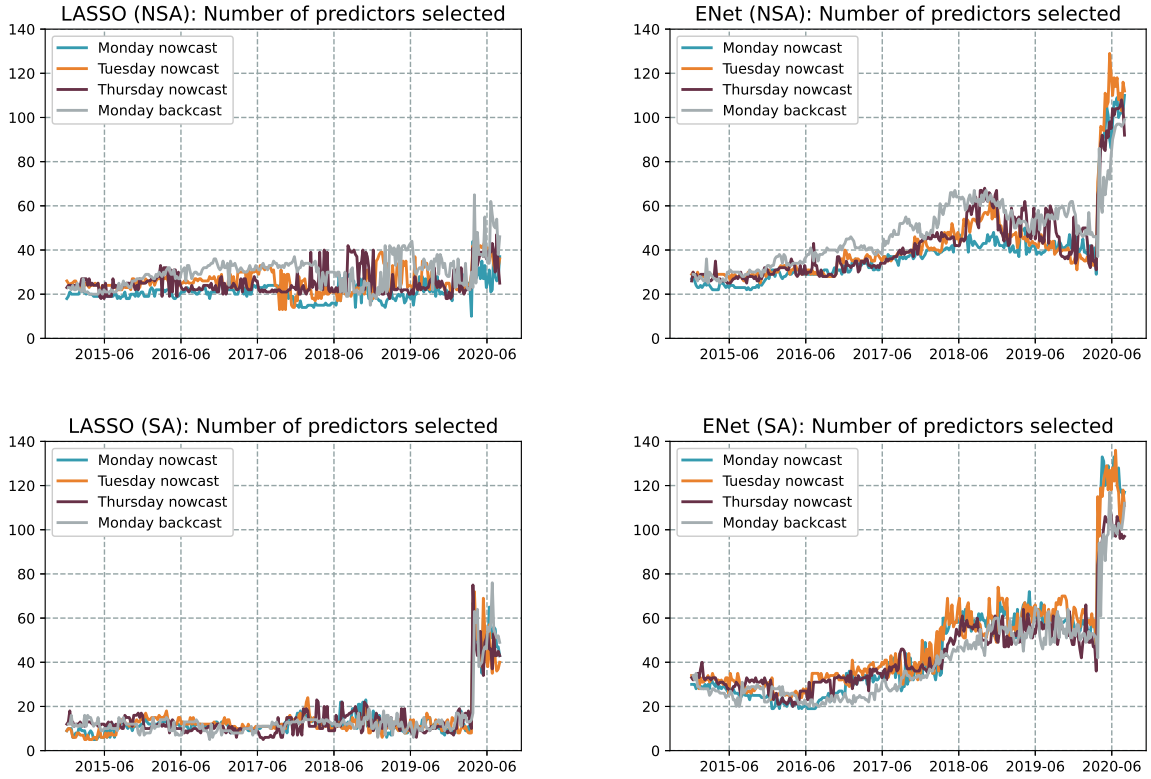


Figure 5: Number of predictors selected by the LASSO and ENet

The figure depicts the number of predictors selected by the LASSO and elastic net (ENet) for rolling-window estimation of linear models used to generate out-of-sample predictions, where each model can include up to 722 predictors. Results are reported for Monday, Tuesday, and Thursday nowcasts and the Monday backcast. Results are reported for non-seasonally adjusted (NSA, upper panels) and seasonally adjusted (SA, lower panels) weekly unemployment insurance initial claims. The out-of-sample period spans the first week of January 2015 through the first week of August 2020.

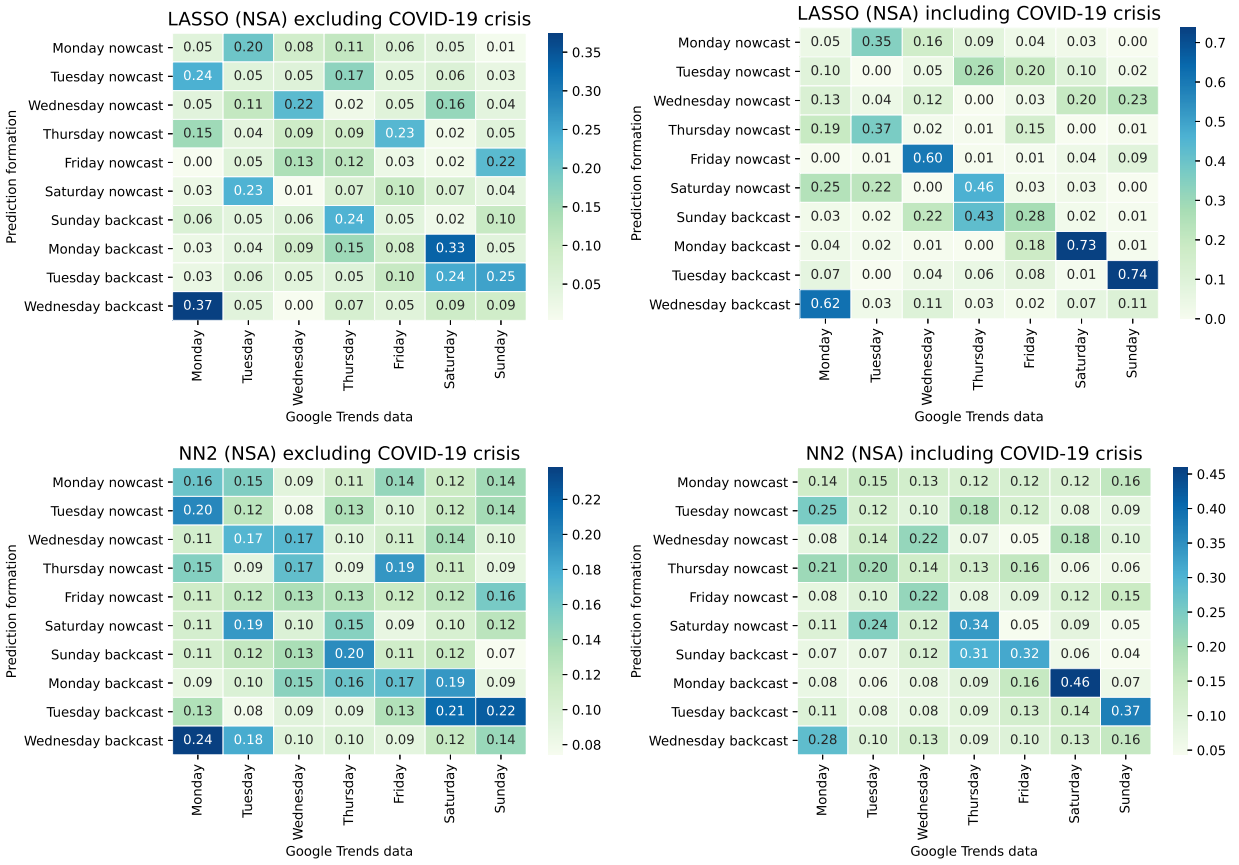


Figure 6: Day-of-the-week effects

The figure depicts heatmaps for joint variable-importance measures of Google Trends terms grouped by the day of the week. The results are for linear models fitted via the LASSO and fitted artificial neural networks with two hidden layers (NN2). The fitted models underpin the now- and backcasts indicated on the vertical axis. The training sample for the left (right) panels ends in the last week of December 2019 (first week of August 2020), thereby excluding (including) the COVID-19 crisis. Results are for non-seasonally adjusted (NSA) weekly unemployment insurance initial claims.

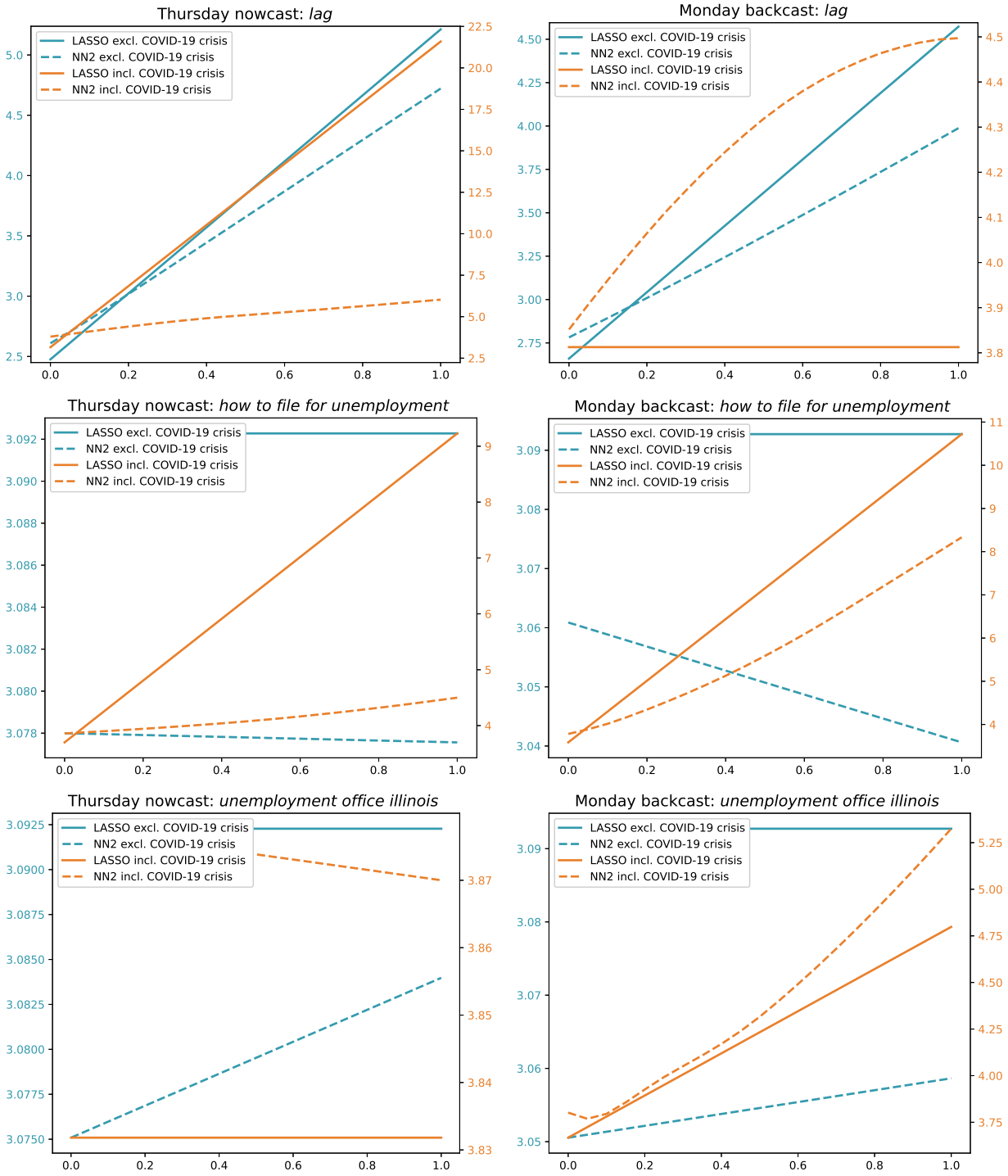


Figure 7: Partial-dependence plots

The figure depicts partial-dependence plots for the autoregressive component (*lag*) and two Google Trends terms (*how to file for unemployment* and *unemployment office illinois*) in linear models fitted via the LASSO and fitted artificial neural networks with two hidden layers (NN2). The values on the horizontal axis are normalized to lie between zero and one. Results are reported for the Thursday nowcast (left panels) and Monday backcast (right panels). The training sample ends in the last week of December 2019 (first week of August 2020), thereby excluding (including) the COVID-19 crisis. Results are for non-seasonally adjusted weekly unemployment insurance initial claims.

Research Papers 2021

2020-06: James G. MacKinnon, Morten Ørregaard Nielsen and Matthew D. Webb: Wild Bootstrap and Asymptotic Inference with Multiway Clustering

2020-07: Javier Hualde and Morten Ørregaard Nielsen: Truncated sum of squares estimation of fractional time series models with deterministic trends

2020-08: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Adaptive Inference in Heteroskedastic Fractional Time Series Models

2020-09: Daniel Borup, Jonas N. Eriksen, Mads M. Kjær and Martin Thyrsgaard: Predicting bond return predictability

2020-10: Alfonso A. Irarrazabal, Lin Ma and Juan Carlos Parra-Alvarez: Optimal Asset Allocation for Commodity Sovereign Wealth Funds

2020-11: Bent Jesper Christensen, Juan Carlos Parra-Alvarez and Rafael Serrano: Optimal control of investment, premium and deductible for a non-life insurance company

2020-12: Anine E. Bolko, Kim Christensen, Mikko S. Pakkanen and Bezirgen Veliyev: Roughness in spot variance? A GMM approach for estimation of fractional log-normal stochastic volatility models using realized measures

2020-13: Morten Ørregaard Nielsen and Antoine L. Noël: To infinity and beyond: Efficient computation of ARCH(∞) models

2020-14: Charlotte Christiansen, Ran Xing and Yue Xu: Origins of Mutual Fund Skill: Market versus Accounting Based Asset Pricing Anomalies

2020-15: Carlos Vladimir Rodríguez-Caballero and J. Eduardo Vera-Valdés: Air pollution and mobility in the Mexico City Metropolitan Area, what drives the COVID-19 death toll?

2020-16: J. Eduardo Vera-Valdés: Temperature Anomalies, Long Memory, and Aggregation

2020-17: Jesús-Adrián Álvarez, Malene Kallestrup-Lamb and Søren Kjærgaard: Linking retirement age to life expectancy does not lessen the demographic implications of unequal lifespans

2020-18: Mikkel Bennedsen, Eric Hillebrand and Siem Jan Koopman: A statistical model of the global carbon budget

2020-19: Eric Hillebrand, Jakob Mikkelsen, Lars Spreng and Giovanni Urga: Exchange Rates and Macroeconomic Fundamentals: Evidence of Instabilities from Time-Varying Factor Loadings

2021-01: Martin M. Andreasen: The New Keynesian Model and Bond Yields

2021-02: Daniel Borup, David E. Rapach and Erik Christian Montes Schütte: Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data