Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Nowcasting unemployment insurance claims in the time of COVID-19
William D. Larson a,1,2,*, Tara M. Sinclair b,1

a Federal Housing Finance Agency, United States of America
b George Washington University, United States of America

ARTICLE INFO
Keywords:
Panel forecasting
Time series forecasting
Forecast evaluation
Structural breaks
Google Trends

ABSTRACT
Near-term forecasts, also called nowcasts, are most challenging but also most important when the economy experiences an abrupt change. In this paper, we explore the performance of models with different information sets and data structures in order to best nowcast US initial unemployment claims in spring of 2020 in the midst of the COVID-19 pandemic. We show that the best model, particularly near the structural break in claims, is a state-level panel model that includes dummy variables to capture the variation in timing of state-of-emergency declarations. Autoregressive models perform poorly at first but catch up relatively quickly. The state-level panel model, exploiting the variation in timing of state-of-emergency declarations, also performs better than models including Google Trends. Our results suggest that in times of structural change there is a bias–variance tradeoff. Early on, simple approaches to exploit relevant information in the cross sectional dimension improve forecasts, but in later periods the efficiency of autoregressive models dominates.

1. Introduction
Simple time series models often forecast well in normal times (see the long literature started by Nelson (1972)), but in the midst of dramatic upheaval, such as in the first few weeks of the response to COVID-19, these models do not adjust quickly to changing conditions (Castle, Clements, & Hendry, 2016). These models exploit a long time series of an aggregate to provide precise, consistent estimates, while typically ignoring information on disaggregate components. We argue that in times of structural change, forecasters should attempt to exploit information within disaggregates, especially when there exists timing variation in the breaks among the disaggregate entities. In the case of the spring of 2020, we find that variation in state-level COVID-19-related state-of-emergency declarations provides useful information for predicting the national quantity of initial unemployment insurance claims.

The number of persons filing initial claims for unemployment insurance (UI) is an important indicator of current US economic conditions (Berge & Jordà, 2011; Lewis, Mertens, & Stock, 2020). UI data are both timely and frequent, and constitute one of a limited number of macroeconomic indicators available at a weekly cadence. As the COVID-19 pandemic came to the US, it became...
an important indicator used to evaluate the state of the economy and the economic toll of the pandemic and associated stay-at-home orders in near-real time. It also itself became a key variable to forecast, to predict the economic costs of the pandemic.  

In this paper we present current-week forecasts (i.e., nowcasts) of the advance estimates of US national initial unemployment claims produced with different information sets and data structures.  

We show that a small T panel model performs well shortly after a structural break if it has relevant information. The variation in emergency declaration date across the states provides this relevant information and outperforms models that include Google Trends data, as well as autoregressive models. We further show that autoregressive models do catch up within a few periods, but that they perform poorly in the crucial weeks directly after the dramatic increase in claims that came in mid-March of 2020.  

Other researchers have recognized the need for different tools to forecast UI claims in the time of COVID-19 beyond simple time series models. Aaronson, Brave, Butters, Sacks, and Seo (2020) use an event-study approach based on hurricanes in order to use Google Trends data to forecast claims. Goldsmith-Pinkham and Sojourner (2020) also use Google Trends at the state level to forecast claims. Our innovation is to use the timing of emergency declarations across the states in a panel framework. We compare our forecasts to those produced using Google Trends in both panel and time series frameworks as well as with autoregressive models. We show that all models miss the initial shock, but that in the early periods the information in the declaration-weeks panel leads to substantially better forecasts. Quickly thereafter, the time series models catch up. Google Trends models are typically second- or third-best in each time period, outperforming autoregressive models initially, but still underperforming declarations dummy variable models.  

Our research also connects to the debate about forecasting the aggregate or aggregating forecasts (see discussion in Castle and Hendry (2010), Hendry and Hubrich (2011), Larson (2015), and Heinisch and Scheufele (2018)). We find support for forecasting the individual states and then aggregating in the case when we have useful variation in the timing of the states, e.g. the emergency declaration dates. This is consistent with the argument from Castle, Fawcett, and Hendry (2011), that we need relevant information available to forecast during breaks. After the usefulness of that information is exhausted, however, simple autoregressive models perform best, with little difference between direct national forecasts and aggregates of state forecasts.  

Research using Big Data for forecasting has emphasized the importance of a large number of time observations for forecasting, but in the face of structural breaks we may not have a long time series with consistent parameters (Bajari, Chernozhukov, Hortaçsu, & Suzuki, 2019). We may, however, have useful information in the panel dimension. Since we have variation in the timing of the emergency declaration by state, we can exploit this information using dummy variables in a panel framework. We show that in this case the cross sectional information does improve forecasts in the periods near the structural break.  

The remainder of our paper is as follows. First, we describe the UI claims, Google Trends, and disaster declaration series. Next, we describe the models used to generate pseudo-real-time forecasts, paying particular attention to information set assumptions regarding variables and timing and the structure of the data. Then, we turn to a discussion of our results, after which we conclude. Detailed model estimates and alternative forecasts can be found in the appendix.  

2. Data

2.1. Unemployment insurance claims

Our target variable is the national total of the advance number of initial UI weekly claims under state programs, not seasonally adjusted (NSA). Fig. 1, panel a reports this variable for each week since January 1998. The previous peak, near the end of the Great Recession, was just under 1 million people filing initial UI claims (NSA). In early 2020, the US economy was relatively strong, with around 215,000 weekly claims each week in February. In response to the COVID-19 pandemic and stay-at-home orders, claims quickly dwarfed the Great Recession high, rising almost 30-fold from this benchmark entry, with claims exceeding 6 million in a single week (panel b). Our analysis is focused on this crisis period and our forecasts begin for the week ending March 14th, 2020. Our objective throughout the paper is to forecast the advance release of the series presented in Fig. 1.  

Initial UI claims are released each Thursday morning at 8:30 a.m. Eastern Time by the US Department of Labor (DOL) and the latest data are the preliminary estimates for the week ending the Saturday five days prior. Each week our data set contains initial claims by state for a total of 52 entities in the panel: 50 US states plus the District of Columbia (DC) and Puerto Rico.  

We refer to all the entities as states for simplicity. The objective of the forecasts will be the advance national totals, which are directly calculated as a sum of the 52 state entities. We focus on NSA numbers throughout our analysis.

3 The insured unemployed in the US is typically a subset of all unemployed, since some jobless do not qualify or apply for benefits. Furthermore, some that qualify for unemployment benefits in the pandemic may not be actively looking for work and therefore may be classified in surveys as out of the labor force rather than unemployed.  

4 Wright (2019) emphasizes the important role of nowcasts in macroeconomic forecasting.  

5 Google Trends has been shown to be useful in other contexts, including forecasting the US unemployment rate (D’Amuri & Marcucci, 2017) and nowcasting GDP, particularly when official data are not available (Ferrara & Simoni, 2019). See Choi and Vian (2012) for discussion of using Google Trends for nowcasting several economic indicators, including unemployment claims.

6 We do not model the US Virgin Islands, which typically has fewer than 100 claims per week.  

7 We are focused on the initial claims under state programs, not the additional claims for the national Pandemic Unemployment Assistance from the CARES Act (information at https://www.dol.gov/coronavirus/unemployment-insurance).
following the recommendation of Rinz (2020), since the multiplicative seasonal adjustment procedure that is used for UI claims is likely misleading in the case of the magnitudes of changes in the sample we are analyzing.\footnote{The use of seasonally adjusted (SA) versus not seasonally adjusted (NSA) numbers has been the subject of some debate in the COVID-19 downturn. Seasonal adjustments to advance weekly UI claims are multiplicative, which we believe to be overestimates of the true seasonal effect. Accordingly, we use NSA numbers throughout.} Our pseudo-real-time data set is constructed using the historical weekly file available on the DOL website through the week ending March 7th, 2020, and then the press release PDF files.\footnote{These files are available at \url{https://oui.doleta.gov/unemploy/DataDownloads.asp} series ETA 539 and \url{https://oui.doleta.gov/unemploy/archive.asp}, respectively.}

All models are estimated on data available at the time the forecast was made. In our exercise, forecasts are assumed to be made immediately following the Thursday release for the next week's release. Unemployment claims data are generally revised once, meaning that the last observation used in each week's estimate will be preliminary data ("advance" estimates in DOL terminology) that is updated the following week.\footnote{For our real-time data set we assume the data are revised only once, so each week we update our data set with the "advance" estimates for the previous week along with revised estimates for the week before that. There was one clear data error where preliminary data for Connecticut for the week ending on May 9th was originally reported as 298,680, but was quickly corrected to 30,000 per the Wall Street Journal: \url{https://www.wsj.com/articles/unemployment-claims-keep-piling-up-in-tri-state-area-11589471382}. We used the corrected number as the preliminary value for Connecticut in this case and adjusted the national values appropriately.}

In our analysis, we develop several forecasting models that exploit different samples of data and variable transformations. For autoregressive models (both state and national), our dependent variable is the natural log of weekly UI claims, and our sample begins in January 1998 in order to exploit a large number of periods, satisfy "Large T" asymptotics, and ensure ergodicity. Augmented Dickey–Fuller tests confirm this series to be stationary despite the rising labor force over the period, with all tests rejecting the null of a unit root at the 1% level.

In models exploiting the panel dimension—"Large N" models—our dependent variable is the normalized weekly US claims for each state relative to the last pre-crisis week, which we identified to be the week ending February 15th. We then reverse this normalization when we aggregate back to the national level to produce the forecast of US national NSA initial unemployment claims. Our panel sample period begins the week ending February 1st.

### 2.2. Emergency declaration dates

Our declaration date analysis exploits the variation of timing in emergency declarations by states so that we can learn from the early states for predicting outcomes for later states, similar to the approach of Liu, Moon, and Schorfheide (2020). The event time is based on the week in which a state of emergency was declared regarding COVID-19 for that state. All states, as well as DC and Puerto Rico, declared a state of emergency within a four-week period in February and March of 2020 (Table 1). Washington state was the first to declare an emergency on February 29th and West Virginia was the last on March 16th. Fig. 2, panels a through d, show normalized state claims grouped by their week of declaration, where 0 on the x-axis for each group is the week that the group declared their emergency. From these figures it is clear that all states experienced a substantial increase in claims after declaring a state of emergency. The last states to declare an emergency already had increases before their declaration week, but continued to experience further increases in the weeks following. It is important to note that we are using declarations of states of emergency, not stay-at-home orders or other restrictions, which, as discussed in Cronin and Evans (2020), explain only a small amount of the change in behavior in response to the pandemic.

#### 2.3. Google Trends

Google Trends data are used as an index of the relative search volume on Google for the keyword “unemployment” for the US and for each of the 50 states plus DC and Puerto Rico.\footnote{Aaronson et al. (2020) also use the keyword “unemployment”. We also considered “file for unemployment” following Goldsmith-Pinkham and Sojourner (2020) with little impact on the results. One interesting finding is that including both keywords helps the national forecast and seems to be a good way to manage sampling error, but we leave deeper exploration of this issue to further research. Borup, Rapach, and Schütte (2020) considered a large number of keywords associated with unemployment and found that their high-dimensional set of Google Trends terms improved nowcasts of UI claims over a benchmark AR model.} The Google Trends API only allows five comparison locations per search, so, following (Goldsmith-Pinkham & Sojourner, 2020), we include California in each of our rounds of data collection and re-normalize each state relative to California. This approach allows us to compare across time and states. In order to get a long daily time series for the national trend, we follow a similar approach, pulling each six-month time period along with the initial six months. Because Google Trends data are based on a sample, we averaged over 10 different samples to reduce the noise (Stephens-Davidowitz & Varian, 2014).

It is important to note that the models that include Google Trends are true nowcasts in the sense that they are using information available into the target week. Google Trends data are available with a delay of approximately 36 h, so for forecasts made on Thursday morning for the data that will be released the following week, we already have Google Trends data through Monday of the reference week. In order to use all information available at the time of our forecast, we create two different Google Trends variables. We include the average of the two days that are available in the current week, which we call $G_t$. We
Fig. 1. National UI Claims. Notes: Data present the advance estimate of U.S. total (50 states plus the District of Columbia and Puerto Rico) weekly initial unemployment insurance claims.

We also include a separate variable that is the average of the previous week, where the dates line up with the latest available UI claims numbers, which we call $G_{t-1}$.

Similar to the UI claims, the Google Trends data are normalized by their average value from the week ending February 15th. Since there are many cases of zeros in particularly for the national Google Trends forecasts that including both $G_t$ and $G_{t-1}$ produces the best forecast in terms of MAE, but with increased variance.

12 We also considered a single variable of the latest 7 days available, but it performed worse in our forecasting models. We also considered excluding $G_{t-1}$, but that produced a worse forecast as well. We find
Google Trends, we first add 1 to all values, and then the Trends are divided by the average of 2/9 through 2/15 in the case of $G$ and by the average value of 2/9 and 2/10 in the case of $\tilde{G}$.

Google Trends data are available from January of 2014. We use the full series for the national time series model. For the panel models we start in February of 2020, consistent with our other panel models.

3. Models

Our forecast target is weekly US national initial unemployment claims with a focus on the first release, which is announced the Thursday of the week following the reference week. The nomenclature for the horizon of our forecast is tricky. Models that do not include Google Trends or disaster emergency declarations data can plausibly be considered one-week-ahead forecasts, since they are based on data through the previous week. However, the UI claims data on the previous week are only released on Thursday morning of the current week, which means we cannot produce the forecasts until well into the current week. Thus if we name the horizon based on when we are able to make the forecast, it would be a current-week forecast, also known as a nowcast. We estimate each of our models with data available on Thursday morning of the reference week. Thus the Google Trends data are available through Monday of that week, and declarations for the current week made after Wednesday are assumed not to have been made at the time of the forecast. Following Banbura, Giannone, Modugno, and Reichlin (2013), we use the term “nowcasts” for all models.

We consider several combinations of data structures and information sets. Data structures considered include US state-level panel models, state time series models that do not exploit the panel dimension, and national time series models that do not exploit any disaggregate information. For the state-level data, we sum up to the

---

13 The national series of normalized claims and Google Trends can be found in the appendix, Figure A.1.
national level for evaluation of the forecast. For information sets, we use single-series autoregressive models, Google Trends at the state and national levels, depending on the model, and dummy variables to convert the event time of the emergency declaration by state into calendar time. In the main results section, we report a total of six different models, where in the case of state-level models we aggregate back to the US national values for the final forecast, which we describe below in greater detail.

Our main model is a panel model with dummy variables representing the distance week \( t \) is from state \( i \)'s emergency declaration date. The variation in this information is what gives our panel good forecasting properties.\(^{14}\) The panel regression is weighted based on covered employment for each state for the week ending March 7th.

\(^{14}\) Once we know the dates of all the emergency declarations, this model is equivalent to a panel structured around the week each state declared an emergency (event time), but we keep it structured around calendar time to ease aggregation and forecasting.

In this model, state differentials in declarations timing give us information about future UI claims behavior. The states are in four different groups based on the timing of their emergency declaration. Washington state was the first to declare an emergency on February 29th, which we classify as a "week 1" declaration. The remaining states all followed suit over the next three weeks (see Table 1) and are given labels of weeks 2 through 4. Our first nowcast was made on March 12th for the week ending March 14th.

The timing of our nowcasting exercise affects the availability of information on declaration dates. To illustrate, for the nowcast for the week ending March 14th, we use all information available up to March 12th, which includes the UI claims release that morning covering the week ending March 7th, and all declarations up to March 11th. All week 1 and week 2 declaration states have dummies set equal to 1 in their respective week. However, for week 3 states, only about half of the states would have dummy variables set equal to 1 for the week ending March 14th in this particular forecast vintage, because declarations after March 11th are not included. All other week 3 and week 4 declaration states have dummy variables set equal
Table 1
COVID-19 disaster declaration dates.
Source: Authors’ assembly based on publicly available news reports.

| Week 1 |     |
|--------|-----|
| Washington | 2/29/2020 |

| Week 2 |     |
|--------|-----|
| California | 3/4/2020 |
| Hawaii | 3/4/2020 |
| Maryland | 3/5/2020 |
| Indiana | 3/6/2020 |
| Kentucky | 3/6/2020 |
| Utah | 3/6/2020 |
| New York | 3/7/2020 |

| Week 3 |     |
|--------|-----|
| Oregon | 3/8/2020 |
| Florida | 3/9/2020 |
| Illinois | 3/9/2020 |
| Iowa | 3/9/2020 |
| New Jersey | 3/9/2020 |
| Ohio | 3/9/2020 |
| Rhode Island | 3/9/2020 |
| Colorado | 3/10/2020 |
| Connecticut | 3/10/2020 |
| Massachusetts | 3/10/2020 |
| Michigan | 3/10/2020 |
| North Carolina | 3/10/2020 |
| Alaska | 3/11/2020 |
| Arizona | 3/11/2020 |
| District of Columbia | 3/11/2020 |
| Louisiana | 3/11/2020 |
| New Mexico | 3/11/2020 |
| Delaware | 3/12/2020 |

| Week 3, Continued |     |
|-------------------|-----|
| Kansas | 3/12/2020 |
| Montana | 3/12/2020 |
| Nevada | 3/12/2020 |
| Puerto Rico | 3/12/2020 |
| Tennessee | 3/12/2020 |
| Virginia | 3/12/2020 |
| Wisconsin | 3/12/2020 |
| Alabama | 3/13/2020 |
| Arkansas | 3/13/2020 |
| Idaho | 3/13/2020 |
| Minnesota | 3/13/2020 |
| Missouri | 3/13/2020 |
| Nebraska | 3/13/2020 |
| New Hampshire | 3/13/2020 |
| North Dakota | 3/13/2020 |
| South Carolina | 3/13/2020 |
| South Dakota | 3/13/2020 |
| Texas | 3/13/2020 |
| Vermont | 3/13/2020 |
| Wyoming | 3/13/2020 |
| Georgia | 3/14/2020 |
| Mississippi | 3/14/2020 |

| Week 4 |     |
|--------|-----|
| Arizona | 3/11/2020 |
| District of Columbia | 3/11/2020 |
| Oklahoma | 3/15/2020 |
| Pennsylvania | 3/16/2020 |
| West Virginia | 3/16/2020 |

Normalized state claims are aggregated to national claims in several steps. First, we undo the normalization by multiplying estimated normalized claims by the February 15th value of state claims. Next, we aggregate to the national level by summing the 52 state estimates. Finally, because the estimator for the normalized data is not unbiased in terms of the national sum, we include a bias-adjustment term $b$ that is equal to the percent error from the last in-sample predicted value. This is akin to intercept correcting a forecast based on the last estimated

to 0 in all periods in this vintage. For the March 21st forecast vintage, the remaining week 3 states would have their dummy variables set equal to 1 for the week ending March 14th, and so on.

In terms of notation, $j$ is how many weeks state $i$ is away from emergency declaration at time $t$. So, for instance, in the week ending March 14th, the dummy variable, $d_{ijt}$ for Washington, whose declaration date is February 29th is 1 for $j = 2$, and 0 otherwise. By the same logic, the dummy variable in the same time period for California (declaration date of March 4th) is 1 for $j = 1$, and 0 otherwise. We also include dummy variables for the two periods in advance of the declaration once the declaration date is known. These leading dummy variables serve two purposes: they absorb any pre-declaration explanatory power; and they serve a “placebo” role in inference related to declaration pre-trends, though that is not a focus of our model design. Thus we estimate $J + 3$ dummy variables, where $J$ is the number of weeks since the week ending February 29th: two leads, one concurrent to the declaration date, and $J$ lags.

$$NormStateClaims_{it} = \beta_0 + \beta_1 \bar{C}_{it} + \beta_2 C_{it-1} + u_{it}$$

We also estimate panel models that include Google Trends data for each state, where $\bar{G}_{it}$ denotes the average of the Sunday and Monday of the current week and $G_{it-1}$ is the average for the entire prior week.$^{15,16}$

Normalized state claims are aggregated to national claims in several steps. First, we undo the normalization by multiplying estimated normalized claims by the February 15th value of state claims. Next, we aggregate to the national level by summing the 52 state estimates. Finally, because the estimator for the normalized data is not unbiased in terms of the national sum, we include a bias-adjustment term $b$ that is equal to the percent error from the last in-sample predicted value. This is akin to intercept correcting a forecast based on the last estimated

$^{15}$ A model with both declaration-week dummy variables and Google Trends performed worse than a model with just declaration-week dummy variables. The results for this model are reported in Table A.5 in the appendix. We encourage further work on model selection in the vein of Hendry and Krolzig (2005).

$^{16}$ A remark is necessary concerning Pandemic Unemployment Assistance (PUA) search intensity and claims. We exclude PUA claims in our analysis, but Google searches for unemployment may be in part driven by PUA-related searches after the CARES Act was signed into law on March 27th. Google searches for the term “PUA” only slowly ramped up and PUA claims only began being tracked in large numbers beginning the week ending April 25th. It is possible there could be a break in the relationship between Google Trends data and regular unemployment claims due to the introduction of PUA during our sample. In this case, the Trends forecasts would over-estimate actual weekly non-PUA claims. We do find evidence of consistent positive bias in the state panel Google Trends forecasts beginning in the week ending April 4th, but the national Google Trends has a negative mean error overall, and a less clear pattern across the weeks, as can be seen in Table 2.
follows, where the modeled variable is normalized US and then summing across the 52 states.

\[
\text{StateClaims}_i = \sum_j \text{NormStateClaims}_j \times \text{StateClaims}_i, \text{Feb15}
\]

where

\[
\hat{b}_t = \frac{\text{USClaims}_{t-1}}{\left( \sum \text{NormStateClaims}_{t-1} \times \text{StateClaims}_i, \text{Feb15} \right)}
\]

The variable \( \hat{b}_t \) consists of lagged values in the data, but is only available once the information set at time \( t \) is available.

The panel models are then compared to state and national time series models. For state-level autoregressive models, we model the natural log of the state claims separately for each state \( i \):

\[
\ln(\text{StateClaims}_i) = \beta_0 + \sum_{j=1}^{p} \phi_j \ln(\text{StateClaims}_{i-j}) + u_t
\]

We aggregate the state claims to the national level, first by exponentiating and adjusting for the variance of the residuals to correct for bias from Jensen’s inequality

\[
\text{StateClaims}_i = \exp \left( \ln(\text{StateClaims}_i) + 1/2 \times \sigma_u^2 \right)
\]

and then summing across the 52 states.

The national-level model with Google Trends is as follows, where the modeled variable is normalized US claims. When calculating the estimated US claims value, we use the same aggregation bias-correction method as the state model, but with national claims instead of state claims.

\[
\text{NormUSClaims}_t = \beta_0 + \beta_1 \tilde{G}_t + \beta_2 G_{t-1} + u_t
\]

The national-level autoregressive model of order \( p \) is estimated in the same way as the states, with the level value calculated using the estimated residual variance to adjust for Jensen’s inequality.

\[
\ln(\text{USClaims}_t) = \beta_0 + \sum_{j=1}^{p} \phi_j \ln(\text{USClaims}_{t-j}) + u_t
\]

### 4. Results

#### 4.1. Weekly summary

We report forecasts for 10 weeks from the week ending on March 14th, 2020, through the week ending May 16th, 2020. The forecasts are depicted in Fig. 3 along with the actual values of the advance data for the national UI claims, which are the target of the forecasts. Forecast percentage errors (defined as forecast minus actual as a share of actual) are reported in Table 2. The models reported line up with the equations in the previous section. Models 1 through 3 use state-level data, while models 4 and 5 use national-level data. Models 1 and 2 use a panel data structure, with our main focus being the panel declarations dummy variable (“declarations DV”) model. The remaining models use a time series data structure.

---

For discussion of aggregation bias, see Baltagi (2008), Goodfriend (1992), and Pesaran and Smith (1995). Our approach is similar to a forecast error intercept correction, except that instead of taking the forecast error from the model used the previous period, we re-estimate the model this period and use the in-sample residual for the last observation, which was the forecast target last period.

We also estimated a state-level AR model with declaration-week dummy variables, but that model performed worse, see model 7 in the appendix, Table A5.

We also estimated a national-level model with Google Trends and AR terms. This model performed better than the Google Trends alone, but not as well as the AR model alone. We report forecasts from the combined model as model 10 in the appendix, Table A5.

Where applicable, state-level forecasts are compared in Table A6.
We estimated several other variants of the models reported, such as including Google Trends in the AR models, panel AR(1) models with Google Trends and state declarations, and robust devices in the vein of Castle, Fawcett, and Hendry (2010) and Castle, Clements, and Hendry (2015), but they typically performed worse than the models reported (see Table A.5).

Table 3 presents our parameter estimates for each vintage week for the state declaration models. We add an additional parameter each week as we move one week further away from the first emergency declaration. Parameter estimates improve from week to week as we get more variation from additional states. For example, if we look at the estimated parameter on the dummy variable when \( j = 2 \) we can follow across the row to see how the parameter varies as we get more information from more states. For the model estimated with data through March 12th for the week ending March 14th, there is only one observation that has a one for this variable: Washington in the latest period. Then the following week, the states in the week 2 group have a one for this variable in the last week, and Washington in the week before, so we get a more precise estimate—and so on for two more weeks until we get as much information as we can get from the states after 4 weeks. Then for week 5, we still have data revisions that affect the parameter estimates, but it is stable after that. A similar pattern appears for all the coefficients on the dummy variables.

For the autoregressive models, both state and national, we report results with \( p = 3 \). We include Google Trends in the panel framework for model 2 and in the national time series framework for model 4.

Focusing on the beginning of the sample, all models perform poorly in the first two weeks. The popular narrative surrounding UI claims highlights the spike for the week ending March 21st, when national claims jumped to a historic high of nearly three million from the previous week of just over 250,000. However, there was also a spike in the week ending March 14th, when claims rose from around 200,000 to 250,000. We can see from Table 2 that in the first week all models miss by double-digit percentage points, and in the second week the absolute forecast errors rise to 76%–92%. In these two periods, no model forecasts substantially better or worse than any other.

In the third week of our forecast analysis, the week ending March 28th, the panel models perform better than the other models, with the state Google Trends model missing by only 10% and the declarations DV model performing second best. The information contained in the declaration week dummy variables was fully populated in this period, with every state having declared a state of emergency by the time of the forecast. The experiences of Washington, California, New York, and some of the earliest-hit states were predictive of patterns exhibited in other states. The timing differentials were key in the forecasting variation in UI claims in states with later declarations.\(^2\)

---

**Table 3**

Model\( [1] \) (Declarations DV Model) Vintage Estimates.

| Parameter | Dependent variable: Normalized state UI claims (StateNormClaims) |
|-----------|---------------------------------------------------------------|
| Forecast Model for Week Ending in Column | | |
| 3/14/2020 | 3/21/2020 | 3/28/2020 | 4/4/2020 | 4/11/2020 | 4/18/2020 | 4/25/2020 | 5/2/2020 | 5/9/2020 | 5/16/2020 |
| \( d_{j=2} \) | −0.0583 | −0.0653* | −0.0653* | −0.0653* | −0.0653* | −0.0653* | −0.0653* | −0.0653* | −0.0653* | −0.0653* |
| \( d_{j=1} \) | 0.0600 | 0.0202 | 0.0218 | 0.0218 | 0.0218 | 0.0218 | 0.0218 | 0.0218 | 0.0218 | 0.0218 |
| \( d_{j=0} \) | −0.0319 | 0.0758** | 1.378 | 1.403 | 1.403 | 1.403 | 1.403 | 1.403 | 1.403 | 1.403 |
| \( d_{j=-1} \) | 0.113*** | 0.170** | 11.39*** | 12.42*** | 12.44*** | 12.44*** | 12.44*** | 12.44*** | 12.44*** | 12.44*** |
| \( d_{j=-2} \) | 1.377*** | 6.122*** | 24.37*** | 24.38*** | 24.42*** | 24.42*** | 24.42*** | 24.42*** | 24.42*** | 24.42*** |
| \( d_{j=-3} \) | 20.50*** | 25.36*** | 32.53*** | 31.77*** | 31.80*** | 31.80*** | 31.80*** | 31.80*** | 31.80*** | 31.80*** |
| \( d_{j=-4} \) | 29.21*** | 26.61*** | 27.62*** | 27.00*** | 27.03*** | 27.03*** | 27.03*** | 27.03*** | 27.03*** | 27.03*** |
| \( d_{j=-5} \) | 27.49*** | 22.62*** | 26.89*** | 26.17*** | 26.23*** | 26.23*** | 26.23*** | 26.23*** | 26.23*** | 26.23*** |
| \( d_{j=-6} \) | 23.25*** | 15.75*** | 21.35*** | 21.01*** | 21.18*** | 21.18*** | 21.18*** | 21.18*** | 21.18*** | 21.18*** |
| \( d_{j=-7} \) | 13.35*** | 11.79*** | 15.24*** | 14.67*** | 14.67*** | 14.67*** | 14.67*** | 14.67*** | 14.67*** | 14.67*** |
| \( d_{j=-8} \) | 22.48*** | 11.90*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** |
| \( d_{j=-9} \) | 16.58*** | 9.294*** | 11.79*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** | 13.48*** |
| \( d_{j=-10} \) | | | | | | | | | | 17.72*** |
| Constant | 1.005*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** | 1.017*** |
| Obs | 260 | 312 | 364 | 416 | 468 | 520 | 572 | 624 | 676 | 728 |
| RMSE | 0.192 | 0.201 | 0.4693 | 0.7364 | 0.8313 | 0.8686 | 10.18 | 11.02 | 10.99 | 10.91 |
| \( R^2 \) | 0.026 | 0.201 | 0.432 | 0.610 | 0.681 | 0.681 | 0.614 | 0.565 | 0.547 | 0.533 |

Notes: This table presents parameters estimated using the column-vintage declarations dummy variable models. These correspond to forecasts made on the date in the column header for the week ending that Saturday (two days following). These models exploit disaster declarations made up to the day prior to the date in the column header. Parameters stabilize after five weeks, due to the presence of four separate weeks of disaster declarations and because some declarations do not occur until after the forecast timing cutoff for that week. In the sixth vintage week, a final revision is made due to revisions to the prior week’s data.

***Standard errors are omitted from the table, with \( p < 0.01 \).

**Standard errors are omitted from the table, with \( p < 0.05 \).

*Standard errors are omitted from the table, with \( p < 0.1 \).

---

\(^2\) We estimated several other variants of the models reported, such as including Google Trends in the AR models, panel AR(1) models with Google Trends and state declarations, and robust devices in the vein of Castle, Fawcett, and Hendry (2010) and Castle, Clements, and Hendry (2015), but they typically performed worse than the models reported (see Table A.5).
This period is crucial because the week ending April 4th was the week of peak claims (and up to this point in the history of the series) with over 6.2 million people filing initial unemployment claims across the country. The percentage forecast error does not fully capture the absolute forecast error in these periods; absolute forecast errors for this period are about 250,000 for the declarations DV model, compared to 1,000,000 to well over 3,000,000 for the alternative models. In the latter case, these errors far exceed the previous high in actual claims before COVID-19 of about 1,000,000 (see the appendix).

In the weeks ending on April 11th and 18th, the relative strength of the forecasting approaches becomes less clear, and it is debatable which forecasting approach is preferred. The trend in the forecast errors for the panel models is clearly pointing to plateauing performance. The state and national AR models, on the other hand, have similar forecast errors but they are showing monotonic improvements in forecasting performance since March 21st. The national Google Trends model continues to forecast poorly in this period.

In the week ending April 25th through the end of the sample on May 16th, the AR models are slightly preferred relative to the panel models, and are much better than the national Google Trends model, though the percent errors are generally smaller across all models and the absolute number of claims is also smaller. Consequently, the cost of choosing an inferior forecast in this period is much lower than in the earlier periods. Absolute forecast errors for the AR models hover between 1% and 6% while the panel models continue to average around 10% to 15%. This corresponds to absolute error differences between the two forecasting approaches of about 100,000 to 200,000.

### 4.2. Discussion

Based on our analysis we can divide our sample into four distinct periods: (1) the first two periods when the crisis is ramping up and all models perform poorly; (2) the next periods when the panel model with declaration DVs stands out; (3) a period of ambiguity where models begin to converge; and (4) at the end when the AR models stand out as best. Fig. 4 illustrates this pattern by zooming in on the forecasts of the declarations DV model and the state autoregressive model. In the periods immediately following the onset of the COVID-19 outbreak, both the level of the claims and the percent errors are larger. Accordingly, over the full sample, the Harvey, Leybourne, and Newbold adjusted Diebold–Mariano (“DM”) test (Diebold & Mariano, 2002; Harvey, Leybourne, & Newbold, 1997) rejects equal predictive accuracy in favor of the declaration-dates panel model when comparing with any of the Google Trends models (models 2 and 4) at the 10% significance level. The AR models also perform worse over the full sample, but not statistically significantly so.

We believe these results have external validity beyond the experience in our specific setting. In general, forecasting UI claims in the time of COVID-19 involves a classic mean–variance tradeoff. Exploiting disaggregate information in the periods immediately following a break acts as a form of costly insurance. Early on, the information gained from the panel model is preferred; in later periods, the inefficiency of estimating individual dummy variable coefficients dominates.

Fig. 5 illustrates this phenomenon at a state level. In the April 4th vintage, the actual values are almost uniformly above the state AR model forecasts. By the May 16th vintage, the bias is mostly gone and is no longer significant. The declarations DV model, on the other hand, is not significantly biased in either period, but it has a larger variance around the actual values, particularly for the week ending May 16th. When we observe these vintage coefficients from the declarations DV models in Table 3, we can see the clear trend in the parameters when looking at descending rows within each column. In the final period model, the coefficients trace out a hump-shaped curve with a tail that is decreasing at a decreasing rate. This is an inefficient way of modeling what is essentially the same type of decay found in an AR...
Fig. 4. State Panel Declarations DV versus State AR Models. Notes: This figure presents the MAEs for two forecasts: [1] (“Declarations DV”) and [3] (“State AR”) from Table 2. The figure shows nearly equal MAEs in the first week (week 1). In weeks 2 and 3, the Panel Declarations DV forecast substantially outperforms the State AR forecast. In weeks 4 and 5, both perform similarly. After week 5, the State AR forecast is consistently more accurate than the Declarations forecast.

(a) Declarations DV, week ending April 4th
\[ \hat{\alpha} = 0.04(0.06), \text{SER} = 0.46 \]

(b) State AR, week ending April 4th
\[ \hat{\alpha} = 0.85(0.05), \text{SER} = 0.39 \]

(c) Declarations DV, week ending May 16th
\[ \hat{\alpha} = 0.10(0.09), \text{SER} = 0.68 \]

(d) State AR, week ending May 16th
\[ \hat{\alpha} = -0.07(0.05), \text{SER} = 0.38 \]

Fig. 5. State-Level Forecasts, Declarations DV versus State AR. Notes: Data present actual and forecast UI claims by state in two weeks: the week ending April 4th and the week ending May 16th. The Declarations DV forecasts are aggregation-bias corrected. Estimated parameters (and standard errors in parentheses) are from the following Holden and Peel (1990) models estimated across states \( i \) within a particular week, where \( f \) is the forecast in the figure panel: \( \ln(\text{StateClaims}_i) - \ln(f_i) = \alpha + u_i \). State-level forecasts are shown in greater detail in Table A.6.
model. Consequently, AR models outperform the panel DV model at the end of the sample.

5. Conclusion

In this paper we produced current-week forecasts (also known as nowcasts) for the national total of state initial unemployment insurance (UI) claims for ten weeks in the midst of the COVID-19 pandemic in the US. We considered different data structures and information sets and compared their performance. We found that in the weeks immediately following the jump in UI claims associated with the COVID-19 crisis, a panel model exploiting the time variation in states declaring a state of emergency performed remarkably well, with the lowest mean absolute error for the full sample across the competing models, and statistically significantly better than models that included Google Trends data available at the time the forecasts were made. Autoregressive models caught up in a few weeks and in the last weeks of the sample had the smallest absolute errors.

Prior research has emphasized the usefulness of simple autoregressive models in normal times, but has recommended much more complicated models for recessions, in order to incorporate sufficient relevant information to identify the change in regime (Chauvet & Potter, 2013). Our findings support the view that simple autoregressive models miss dramatic changes, but we showed that it is possible in certain instances to exploit panel information if there is information on differences in timing in the cross section. This allows us to still use a simple model for forecasting in a time of structural change. Our analysis also emphasizes that autoregressive models are again remarkably useful just a few periods after a break, consistent with the findings of Schorfheide and Song (2020), Primiceri and Tambalotti (2020), and Lenza and Primiceri (2020).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2021.01.001.

References

Aaronson, D., Brave, S. A., Butters, R., Sacks, D. W., & Seo, B. (2020). Using the eye of the storm to predict the wave of Covid-19 UI claims. Available at SSRN 3561298.

Bajari, P., Chernozukhov, V., Hortaçsu, A., & Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. In AEA papers and proceedings, vol. 109 (pp. 33–37).

Baqaee, B. H. (2008). Forecasting with panel data. Journal of Forecasting, 27(2), 153–173.

Banbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Nowcasting and the real-time data flow. In G. Elliott, & A. Timmermann (Eds.), Handbook of economic forecasting: Vol. 2, Handbook of economic forecasting (pp. 195–237). Elsevier. (Chapter 4).

Berge, T. J., & Jordà, Ò. (2011). Evaluating the classification of economic activity into recessions and expansions. American Economic Journal: Macroeconomics, 3(2), 246–277.

Borup, D., Rapach, D., & Schütté, E. C. M. (2020). Now- and backcasting initial claims with high-dimensional daily internet search-volume data. Available at SSRN 3690832.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015). Robust approaches to forecasting. International Journal of Forecasting, 31, 99–112.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2016). An overview of forecasting facing breaks. Journal of Business Cycle Research, 12, 3–23.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2010). Forecasting with equilibrium-correction models during structural breaks. Journal of Econometrics, 158(1), 25–36.

Castle, J. L., Fawcett, N. W., & Hendry, D. F. (2011). Forecasting breaks and forecasting during breaks. University of Oxford Discussion Paper Series.

Castle, J. L., & Hendry, D. F. (2010). Nowcasting from disaggregates in the face of location shifts. Journal of Forecasting, 29(1–2), 200–214.

Chauvet, M., & Potter, S. (2013). Forecasting output. In Handbook of Economic Review, Vol. 2 (pp. 141–194). Elsevier.

Choi, H., & Varian, H. (2012). Predicting the present with google trends. Economic Record, 88, 2–9.

Cronin, C., & Evans, W. N. (2020). Private precaution and public restrictions: What drives social distancing and industry foot traffic in the Covid-19 era?: NBER Working Paper No. w27531.

D’Amuri, F., & Marcucci, J. (2017). The predictive power of google searches in forecasting US unemployment. International Journal of Forecasting, 33(4), 801–816.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1), 134–144.

Ferrara, L., & Simoni, A. (2019). When are Google data useful to nowcast GDP? An approach via pre-selection and shrinkage: Banque de France Working Paper April 2019, WP.

Goldsmith-Pinkham, P., & Sojourner, A. (2020). Predicting initial unemployment insurance claims using google trends. https://paulgp.github.io/GoogleTrendsUINowcast/google_trends_UI.html.

Goodfriend, M. (1992). Information-aggregation bias. The American Economic Review, 82(5), 888–898.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281–291.

Heinisch, K., & Scheufele, R. (2018). Bottom-up or direct? Forecasting german GDP in a data-rich environment. Empirical Economics, 54(2), 705–745.

Hendry, D. F., & Hubrich, K. (2011). Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate. Journal of Business & Economic Statistics, 29(2), 216–227.

Hendry, D. F., & Krolzig, H.-M. (2005). The properties of automatic GETS modelling. The Economic Journal, 115(502), C32–C61.

Holden, K., & Peel, D. A. (1990). On testing for unbiasedness and efficiency of forecasts. The Manchester School, 58(2), 120–127.

Larson, J. W. (2015). Forecasting on aggregate in the presence of structural breaks in the disaggregates: Discussion paper. The George Washington University, Department of Economics, Research Program on Forecasting.

Lenza, M., & Primiceri, G. E. (2020). How to estimate a VAR after March 2020, Manuscript, Northwestern University.

Lewis, D., Mertens, H. J., & Rock, J. J. (2020). US economic activity during the early weeks of the SARS-CoV-2 outbreak: Discussion paper. National Bureau of Economic Research.
Liu, L., Moon, H. R., & Schorfheide, F. (2020). Panel forecasts of country-level Covid-19 infections: CEPR discussion paper no. DP14790.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the US economy. The American Economic Review, 62(5), 902–917.

Neumark, D., Zhang, J., & Ciccarella, S. (2008). The effects of Wal-Mart on local labor markets. Journal of Urban Economics, 63(2), 405–430.

Pesaran, M. H., & Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics, 68(1), 79–113.

Pretis, F., Schneider, L., Smerdon, J. E., & Hendry, D. F. (2016). Detecting volcanic eruptions in temperature reconstructions by designed break-indicator saturation. Journal of Economic Surveys, 30(3), 403–429.

Primiceri, G. E., & Tambalotti, A. (2020). Macroeconomic forecasting in the time of COVID-19. Manuscript, Northwestern University.

Rinz, K. (2020). Understanding unemployment insurance claims and other labor market data during the COVID-19. https://kevinrinz.github.io/covid19_labordata.pdf.

Schorfheide, F., & Song, D. (2020). Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic: Discussion paper, Federal Reserve Bank of Philadelphia.

Stephens-Davidowitz, S., & Varian, H. (2014). A hands-on guide to google data.

Wright, J. H. (2019). Some observations on forecasting and policy. International Journal of Forecasting, 35(3), 1186–1192.