Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data*

Juan Antolín-Díaz
London Business School

Thomas Drechsel
University of Maryland

Ivan Petrella
Warwick Business School
CEPR

August 7, 2020

*Preliminary draft – latest version available **here***

Abstract

The assessment of macroeconomic conditions in real time is challenging. Dynamic factor models, which summarize the comovement across many macroeconomic time series as driven by a small number of shocks, have become the workhorse tool for 'nowcasting' activity. This paper develops a novel dynamic factor model that explicitly captures three salient features of modern business cycles: low frequency movements in long-run growth and volatility, lead-lag patterns in the responses of variables to common shocks, and fat-tailed outliers. We use real-time unrevised data for the last two decades and cloud computing technology to conduct an out-of-sample evaluation exercise of the model. The exercise demonstrates the importance of considering these features for forecasting and probability assessment of economic conditions. In an application to the COVID-19 recession, we develop a method to incorporate newly available high-frequency data. The use of such alternative data is essential to track the downturn in activity, but a careful econometric specification is just as important.

Keywords: Nowcasting; Dynamic factor models; Real-time data; Bayesian Methods; Fat Tails.
JEL Classification Numbers: E32, E23, O47, C32, E01.

*We thank seminar participants at the Bank of England, the Bank of Spain, the Barcelona Summer Forum, the CFE-ERCIM in London, the DC Forecasting Seminar at George Washington University, Fulcrum Asset Management, the IHS in Vienna and the World Congress of the Econometric Society. Author details: Antolín-Díaz: London Business School, 29 Sussex Place, London NW1 4SA, UK; E-Mail: jantolindiaz@london.edu. Drechsel: Department of Economics, University of Maryland, Tydings Hall, College Park, MD 20742, USA; E-Mail: drechsel@umd.edu. Petrella: Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK. E-Mail: Ivan.Petrella@wbs.ac.uk.

1 Introduction

The real-time assessment of macroeconomic conditions is an important responsibility of economists, but remains a challenging task due to the idiosyncrasies of macroeconomic data collection and publication: the most important indicators, such as Gross Domestic Product (GDP), are published on a quarterly basis and with considerable delay, while related but noisy indicators are available in a more timely fashion; most macroeconomic series are revised over time, and many are plagued with missing observations or have become available only very recently. Moreover, as the events of the first two decades of this century have made clear, secular trends, evolving uncertainty and large shocks are pervasive features of modern macroeconomic fluctuations.

This paper advances macroeconomic "nowcasting" by exploiting recent advances in Bayesian computational methods. Our approach gives a prominent role to what Sims (2012) calls "recurrent phenomena" of macroeconomic time series that the literature has traditionally treated as "nuisance parameters to be worked around": low-frequency changes in trends, sustained periods of persistently high or low volatility and large outliers. We propose an econometric framework that enables exact finite-sample inference in high-dimensional, non-linear and non-Gaussian settings. In a comprehensive forecasting evaluation exercise, we conclude that explicitly modeling these features significantly improves the ability to track economic activity in real time. Our results have implications for modelling of macroeconomic time series that go beyond forecasting.

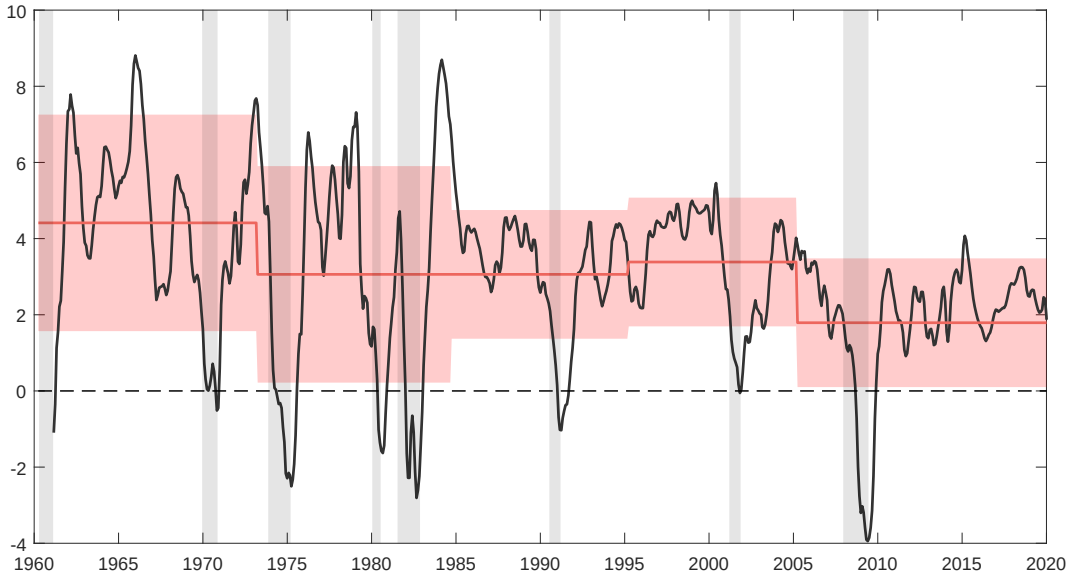
After the first versions of this paper were circulated, the COVID-19 pandemic and associated recession began to unfold. We use this event as an out-of-sample case study to understand the importance of non-linearities and large shocks in understanding the evolution of economic activity as the crisis unfolds in the United States. In response to the pandemic a number of high-frequency indicators have appeared −including credit card transactions, payroll data, or mobility statistics− that can give a timely reading of economic conditions but whose available history is sometimes limited to a few months. We propose a method to incorporate such indicators into our econometric model, and quantify their contribution to the timeliness of the nowcasts. We conclude that while high-frequency data is essential for tracking the COVID-19 recession and recovery, a careful econometric specification is just as important.

Our model is a Bayesian dynamic factor model (DFM) that builds on Geweke (1977), Sargent and Sims (1977), Stock and Watson (1989), and Giannone, Reichlin, and Small (2008), modified to account for low-frequency variation in the mean and variance of the series, heterogeneity in the impulse responses of variables to common shocks, and fat-tailed observations. Figure 1 illustrates the importance of these features for US macroeconomic time series. As documented by Antolin-Diaz, Drechsel, and Petrella (2017), modelling slow-moving changes in the mean and variance of activity are crucial to capture economic phenomena such as the recent productivity slowdown (Fernald, 2014), the Great Moderation (McConnell and Perez-Quiros, 2000), and cyclical variation in uncertainty (Jurado et al., 2014) (Panel a). Moreover, the most common DFM specification implicitly assumes that the impulse responses of different variables to aggregate shocks are proportional to each other. This restrictive assumption of homogeneous dynamics rules out patterns of leading and lagging indicators observed. Panel (b) makes clear that this is a salient feature of the data. Allowing for such heterogeneous dynamics will affect the shape of the recoveries implied by DFM forecasts after an aggregate shock. Finally, panel (c) illustrates how large, transitory one-off innovations, usually caused by tax changes, strikes or natural disasters, are endemic to macroeconomic time series, especially at monthly frequencies. While the literature has typically worked around these observations with ad-hoc methods, we show that a careful treatment of outliers is important when dealing with datasets that include the events surrounding the COVID-19 crisis.

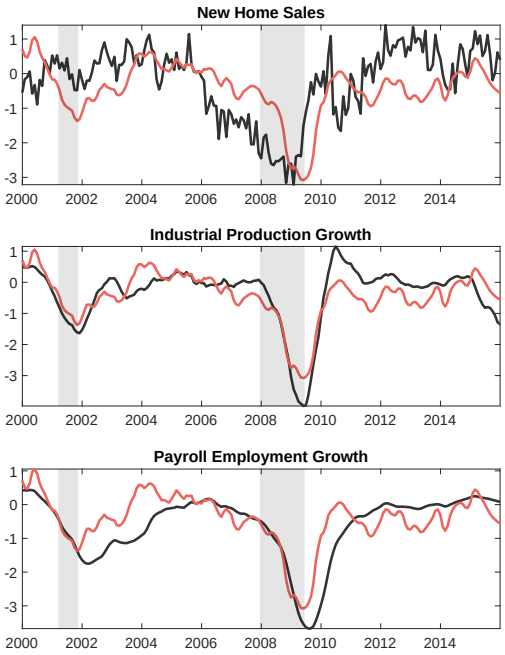
We put our modeling innovations to the test in a comprehensive out-of-sample evaluation exercise. Using a new fully real-time database for the United States, we mimic the exercise of a forecaster that updates her information set in real time every day from January 2000 to December 2019. The sheer scale of such as exercise would have been infeasible just a few years ago. We enable it by developing an efficient hierarchical Monte Carlo Markov Chain (MCMC) algorithm that is fast to estimate the model despite its complexity, and through the use of modern cloud computing techniques. We find that our model delivers a sizeable improvement in tracking and nowcasting real GDP. Relative to a benchmark model without our methodological contributions, both point and density forecasting metrics are statistically significantly and economically meaningfully improved. This is the case across forecasting horizons and throughout the evaluation window. Each of the model components we introduce contributes to the improvement individually.

Figure 1: SALIENT FEATURES OF THE MACROECONOMIC DATA FLOW

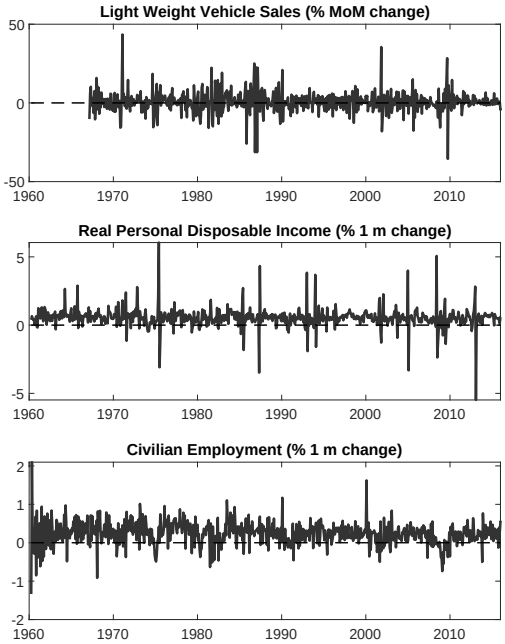
(a) Time-varying long-run growth and volatility in US GDP



(b) Heterogeneous lead/lag dynamics



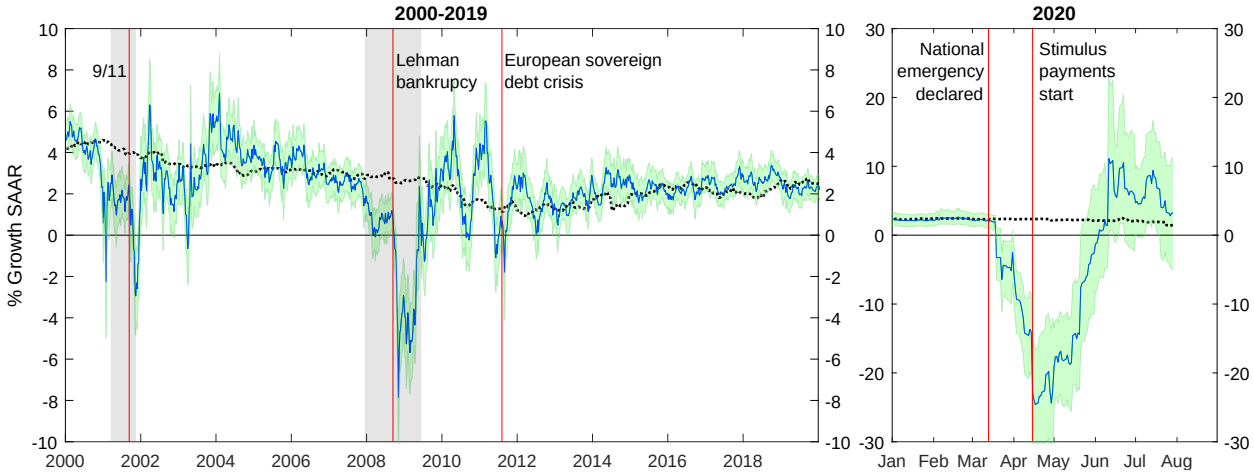
(c) Outliers in macroeconomic indicators



Notes. Panel (a) plots four-quarter US real GDP growth as the black line. The red line and shaded area display the average and standard deviation of GDP growth over selected subsamples, which both reveal meaningful time-variation. Panel (b) displays the twelve-month growth rate of selected indicators of economic activity (blue) together with the four-quarter growth of real GDP (red) over the period 2000-2015. This illustrates how the business-cycle comovement among macroeconomic variables features heterogeneous patterns of leading and lagged dynamics. Panel (c) presents raw data series for selected indicators of economic activity. The variation in the raw series highlights the presence and different nature of outliers in macroeconomic time series. In all panels, gray shaded areas indicate NBER recessions.

We also apply the model out-of-sample to track economic activity in the period surrounding the COVID-19 pandemic. Figure 2 shows our model's daily measure of real economic activity for the periods 2000-2019 and January-June 2020. The dramatic speed and size of the recession poses unique challenges to macroeconometric models. Therefore, the lockdown of 2020 provides a case study to understand their strengths and weaknesses.

Figure 2: DAILY REAL-TIME ESTIMATE OF QUARTER-ON-QUARTER REAL GDP GROWTH



Notes. Daily estimate of US real GDP growth (blue line) with associated uncertainty (green shaded areas). This is computed from our Bayesian DFM described in detail in the text. The left panel plots this estimate from 2000 through to the end of 2019, while the right panel separately presents the period January-June 2020. Note the different scale of the two panels, which indicates the dramatic magnitude of the COVID-19 recession. The vertical red lines indicate significant events. Gray shaded areas indicate NBER recessions.

Several insights arise from this out-of-sample case study. First, we show that models without t-distributed outliers behave erratically and require ad-hoc censoring of COVID-19 observations for the estimation to converge. Instead, our model with fat-tailed innovations produces stable estimates throughout and is able to capture features like the strong rebound of economic activity during the partial re-opening of the economy. Second, we examine ways to incorporate more timely information from novel indicators which are published at daily frequency but have an extremely short history available. We achieve this by conditioning the last observations of closely-related traditional series, for which a long history is available, with the value implied by the more timely novel series. For example, we impose that daily credit card spending data is informative in an analogous way to consumption data, or cell phone mobility data in an analogous way to vehicle miles traveled. By incorporating the high-frequency indicators within the familiar DFM framework,

we show that they contribute to a more timely reading of economic conditions. Nevertheless, different model specifications can lead to qualitatively different forecasts of future activity, even across models that share the same information set. This highlights that a careful econometric specification remains key to tracking economic activity in real time. Our preferred specification detected a strong rebound of activity in May and June 2020.

Contribution to the literature. This paper contributes to the literature that applies DFMs to track developments in economic activity in real time (for applications of DFMs to nowcasting, see Giannone et al., 2008 and Aruoba et al., 2009; for a review of the literature see, e.g., Banbura et al., 2010). Antolin-Diaz et al. (2017) and Doz et al. (2020) highlight the importance of allowing for both shifts in long run growth and changes in volatility.[1] D'Agostino et al. (2015) stress the importance of accounting for the asynchronous nature of macroeconomic data at high frequency within a DFM.[2] The focus of the present paper is to explore the contribution of salient features of modern macroeconomic fluctuations such as secular trends, and large shocks, to the real-time assessment of economic activity. Moreover, in contrast to most of the literature we give emphasis to probabilistic prediction and highlight the ability of the model to produce well calibrated density forecasts. Therefore, our model can be reasonably used to assess the underlying downward risks to economic activity in real time.[3]

Our work also contributes to the efforts to understand the impact of the COVID-19 crisis on macroeconomic time series, modelling and forecasting. While the literature is rapidly evolving, examples include Primiceri and Tambalotti (2020); Lenza and Primiceri (2020), Schorfheide et al. (2020), and Cimadomo et al. (2020) all of which discuss the challenges of estimation and forecasting using vector autoregression models during the pandemic period. Diebold (2020) studies the performance of the Aruoba et al. (2009) model during the pandemic. Relative to these papers, our key contribution is to understand how a model that explicitly features changes in volatility and large outliers fares during this period. Lewis et al. (2020) highlight the usefulness of using weekly

[1] Marcellino et al. (2014) were the first to introduce time varying volatility in a DFM framework.

[2] Andreini et al. (2020) relax the underlying linearity assumption of the DFM. While our model remains conditionally linear, we show that the introduction of a Student-t distributed outlier component is such that the information contained in the data is effectively treated nonlinearly, heavily discounting large idiosyncratic outliers.

[3] See Carriero et al., 2020 for an alternative approach.

data for macroeconomic monitoring during this period, but to the best of our knowledge we are the first to integrate novel daily data series such as credit card expenditures and mobility statistics, with sometimes just a few months of history, into the DFM framework.

Structure of the paper. Section 2 introduces the econometric framework. This includes the structure of the DFM, as well as the estimation algorithm and the construction of our real-time vintage data base. Section 3 illustrates how major challenges in nowcasting are addressed with the novel components of the model. The results for the out-of-sample evaluation exercise are presented in Section 4. Section 5 presents to application to the COVID-19 recession. Section 6 concludes.

2 Econometric framework

DFMs capture the idea that a few aggregate shocks drive most of the comovement between macroeconomic series, which allows to combine the information in many series to extract an indicator of business cycle conditions. Our framework builds on the DFM with time-varying long-run growth and stochastic volatility proposed by Antolin-Diaz, Drechsel, and Petrella (2017), extended to allow for heterogeneous responses to the common factors as well as fat-tailed outliers. Much of the existing DFM literature employs frequentist estimation and inference.[4] We instead take a Bayesian approach to explicitly model features of the data that are typically treated as "nuisance parameters to be worked around" (Sims, 2012).

2.1 The model

Let $\mathbf{y}_t$ be an $n \times 1$ vector of observable macroeconomic time series. A small number, $k << n$, of latent common factors, $\mathbf{f}_t$ is assumed to capture the majority of the comovement between the growth rates of the series. Moreover, the raw data also displays outliers, denoted $\mathbf{o}_t$. Formally,

$$\Delta(\mathbf{y}_t - \mathbf{o}_t) = \mathbf{c}_t + \mathbf{\Lambda}(\mathbf{L})\mathbf{f}_t + \mathbf{u}_t, \tag{1}$$

[4]Note that principal component approaches and he EM algorithm (Doz et al. 2012) have been shown to be robust to non-Gaussianity and other sources of misspecification for a large numnber of series. The asynchronous release of macroeconomic indicators in real time −known as the "ragged edge" problem− implies that N may be small early in the nowcasting period, precisely when accuracy needed. With a Bayesian approach we do not rely on asymptotic results.

where $\mathbf{\Lambda}(\mathbf{L})$ is a matrix polynomial of order m in the lag operator containing the loadings on the contemporaneous and lagged common factors, and $\mathbf{u}_t$ is a vector of idiosyncratic components. The first difference operator, Δ, makes clear that it is the *level* of the variables that displays outliers, while the factor structure is present in the growth rates. The former choice mimics the standard practice, where typically the statistical agency adjusts for outliers on the level of the series, and the data is differenced when appropriate to achieve stationarity before estimating the DFM.[5] It is also convenient because, many time series related to real economic activity feature large one-off innovations to the level, such as strikes and weather related disturbances, that are purely transitory in nature and the series returns to its original level once their effect dissipates.

Following Antolin-Diaz et al. (2017), low frequency shifts in the long-run growth rate of $\mathbf{y}_t$ are captured by time-variation in $\mathbf{c}_t$. In principle one could allow time-varying intercepts in all or a subset of the variables in the system, and common time-variation could be shared between different series. For instance, balanced-growth theory would suggest that the long-run growth component is shared between output and consumption. $\mathbf{c}_t$ is therefore flexibly specified as

$$\mathbf{c}_t = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{c} \end{bmatrix} \begin{bmatrix} \mathbf{a}_t \\ 1 \end{bmatrix}, \tag{2}$$

where $\mathbf{a}_t$ is an $r \times 1$ vector of time-varying means, $\mathbf{B}$ is an $m \times r$ matrix which governs how the time-variation affects the corresponding observables, and $\mathbf{c}$ is an $(n-m) \times 1$ vector of constants.

The relevant laws of motion are specified as

$$(\mathbf{I} - \mathbf{\Phi}(\mathbf{L}))\mathbf{f}_t = \mathbf{\Sigma}_{\varepsilon_t}\boldsymbol{\varepsilon}_t, \tag{3}$$

$$(1 - \rho_i(L))u_{i,t} = \sigma_{\eta_{i,t}}\eta_{i,t}, \quad i = 1, \ldots, n \tag{4}$$

$$o_{i,t} \overset{iid}{\sim} t(v_{o,i}), \quad i = 1, \ldots, n \tag{5}$$

where $\mathbf{\Phi}(\mathbf{L})$ and $\rho_i(L)$ denote polynomials in the lag operator of orders p and q. $\mathbf{I}$ is an identity matrix of conformable size. The idiosyncratic components are cross-sectionally orthogonal and

[5] In practice, some variables, such as business surveys, are already stationary in levels, so the difference operator would not apply to them.

uncorrelated with the common factor at all horizons, i.e. $\boldsymbol{\varepsilon}_t \overset{iid}{\sim} N(\mathbf{0}_{2\times 2}, \mathbf{I}_2)$ and $\eta_{i,t} \overset{iid}{\sim} N(0,1)$. Moreover, the outlier components are independent from the factor and idiosyncratic innovations. The model's time-varying parameters are governed by driftless random walks:

$$a_{j,t} = a_{j,t-1} + v_{a_{j,t}}, \qquad v_{a_{j,t}} \overset{iid}{\sim} N(0, \omega^2_{a,j}) \quad j = 1, \ldots, r \tag{6}$$

$$\log \sigma_{\varepsilon_t} = \log \sigma_{\varepsilon_{t-1}} + v_{\varepsilon,t}, \qquad v_{\varepsilon,t} \overset{iid}{\sim} N(0, \omega^2_{\varepsilon}) \tag{7}$$

$$\log \sigma_{\eta_{i,t}} = \log \sigma_{\eta_{i,t-1}} + v_{\eta_{i,t}}, \qquad v_{\eta_{i,t}} \overset{iid}{\sim} N(0, \omega^2_{\eta,i}) \quad i = 1, \ldots, n \tag{8}$$

where $a_{j,t}$ are the r time-varying elements in $\mathbf{a}_t$, and σ_{ε_t} and $\sigma_{\eta_{i,t}}$ capture the SV of the innovations to factor and idiosyncratic components. The random walk specification follows Primiceri (2005). Finally, the outliers are modeled as independent additive Student-t innovations, with the degrees of freedom, $v_{o,i}$, to be estimated jointly with the other parameters of the model.

2.2 Dealing with mixed frequencies and missing data

The model is specified at monthly frequency, while the (observed) growth rates of quarterly variables, x^q_t, are related to the (unobserved) monthly growth rate x^m_t as

$$x^q_t = \frac{1}{3}x^m_t + \frac{2}{3}x^m_{t-1} + x^m_{t-2} + \frac{2}{3}x^m_{t-3} + \frac{1}{3}x^m_{t-4}, \tag{9}$$

where every third observation of x^q_t is observed (see Mariano and Murasawa, 2003). This insight reduces the presence of mixed frequencies to a problem of missing data in a monthly model. Besides mixed frequencies, additional sources of missing data include the "ragged edge" at the sample end coming from the non-synchronicity of data releases; missing data at the beginning of the sample for more recent series; and data collection errors. Our Bayesian method exploits the state space representation of the DFM and jointly estimates the latent factors, the parameters, and the missing data points using the Kalman filter (see Durbin and Koopman, 2012, for a textbook treatment).

2.3 Priors and model settings

The number of lags in $\mathbf{\Lambda}(\mathbf{L})$, $\phi(L)$, and $\rho(L)$ is set to $m = 1$, $p = 2$, and $q = 2$. We found that $m = 1$ is enough to allow for rich heterogeneity in the dynamics and increasing m does not improve in-sample fit. By setting $p = q = 2$, which follows Stock and Watson (1989), the model allows for the hump-shaped responses to aggregate shocks commonly thought to characterize macroeconomic time series. One of the advantages of the Bayesian approach is that an a-priori preference for simpler models can be naturally encoded by shrinking the parameters towards a more parsimonious specification. We follow the long tradition in economics of applying stronger shrinkage to more distant lags initiated by Doan et al. (1986). "Minnesota"-style priors are applied to the coefficients in $\mathbf{\Lambda}(L)$, $\phi(L)$ and $\rho_i(L)$. For the autoregressive coefficients of the factor, $\phi(L)$, the prior mean is set to 0.9 for the first lag, and to zero in subsequent lags. This reflects a belief that the common factor captures a highly persistent but stationary business cycle process. For the factor loadings, $\mathbf{\Lambda}(L)$, the prior mean is set to 1 for the first lag, and to zero in subsequent lags. This shrinks the model towards the factor being the cross sectional average of the variables, see D'Agostino et al. (2015). For the autoregressive coefficients of the idiosyncratic components, $\rho_i(L)$ the prior is set to zero for all lags, thus shrinking the model towards a model with no serial correlation in $u_{i,t}$. In all cases, the variance on the priors is set to $\frac{\gamma}{h^2}$, where γ is a parameter governing the tightness of the prior, and h is equal to the lag number of each coefficient, ranging $1:p$, $1:q$ and $1:s+1$. We set $\gamma = 0.2$, the reference value used in the Bayesian VAR literature. For the variances of the innovations to the time-varying parameters ω_a^2, ω_ε^2 and $\omega_{\eta,i}^2$ we also use priors to shrink these variances towards zero, i.e. towards the standard DFM which excludes time-varying long-run GDP growth and SV. In particular, for ω_a^2 we set an inverse gamma prior with one degree of freedom and scale equal to 0.001. For ω_ϵ^2 and $\omega_{\eta,i}^2$ we set an inverse gamma prior with one degree of freedom and scale equal to 0.0001. Further discussion is provided in Antolin-Diaz et al. (2017).

2.4 Estimation algorithm

We build a highly efficient Gibbs sampler algorithm to sample from the joint posterior distribution of the parameters and latent objects, adapting ideas from Kim and Nelson (1999), Bai and Wang (2015),

and Moench et al. (2013). The SV step is adapted from Kim et al. (1998) and we draw the Student-t distributed innovations following Jacquier et al. (2004). The standard approach for writing the state-space of the model in (1)-(9) would involve including idiosyncratic and outlier terms as additional state variables (see, e.g. Banbura and Modugno, 2014). This is problematic, as computation time increases with the number of states, which in turn would depend on n, the number of variables. Our contribution is the insight that a hierarchical re-writing of the algorithm avoids extremely large state-spaces. Together with additional computational gains already proposed by Antolin-Diaz et al. (2017), the estimation is made feasible and extremely fast. A sketch of the algorithm is provided below, and full details are provided in Appendix A.

Algorithm 1. *This algorithm draws from the posterior distribution of the unobserved components and parameters of the model described in section 2.1*

1. *Initialize the parameters of the model as well as the stochastic volatility processes at their prior means; The latent components* $\mathbf{c}_t, \mathbf{s}_t, \mathbf{o}_t$, *and* $\mathbf{f}_t$, *are initialized by running the Kalman filter and smoother once conditional on the initialized parameters.*
2. *For each variable,* $i = 1, \ldots, N$*:*
   - 2.1. *Compute* $\Delta y_{i,t} - c_{i,t} - \lambda_i(L) f_t = u_{i,t} + \Delta s_{i,t} + \Delta o_{i,t}$.
   - 2.2. *Use the Kalman filter and simulation smoother to independently draw the outlier components. Draw the associated parameters* $\sigma_{\varepsilon_s,t}$, $v_{o,i}$.
   - 2.3. *Compute the outlier adjusted variable,* $\Delta y_{i,t}^{SA} = \Delta(y_{i,t} - s_{i,t} - o_{i,t}) = c_{i,t} + \lambda_i(L) f_t + u_{i,t}$.
3. *Conditional on the outlier adjusted variables:*
   - 3.1. *Use the Kalman filter and simulation smoother to draw* $\{\mathbf{c}_t, \mathbf{f}_t\}$.
   - 3.2. *Conditional on the factors and long-run trends, draw the remaining parameters of the model,* $\mathbf{\Lambda}(L)$, $\phi(L)$ *and* $\rho_i(L)$, *as well as the stochastic volatility processes, the innovations to the time-varying parameters,* ω_a^2, ω_ε^2 *and* $\omega_{\eta,i}^2$.
4. *Go back to Step 2 until convergence has been achieved.*

Several points are worth noting. First, the algorithm iterates between two state spaces: a univariate one in Step 2, which performs outlier adjustment, and a multivariate one in Step 3, which estimates a DFM on the outlier adjusted variables. It therefore mimics the usual practice of using independently outlier adjusted data in the factor model, but incorporates the uncertainty

inherent in the outlier adjustment process, which is typically disregarded. Second, the univariate state space in Step 2 is independent across variables, which means that it can be run in parallel using multi-core processors.[6] Finally, by incorporating the mixed-frequency measurement into Step 2 only, the maximum size of the state-space in Step 3 is limited if the system is re-written in terms of quasi-differences, avoiding the inclusion of idiosyncratic components as state variables.

2.5 Variable selection

Our DFM includes variables measuring real activity, excluding prices, monetary and financial variables, and is specified with a single common factor to capture real activity.[7] We include all of the available surveys of consumer and business sentiment. The timeliness of these *soft* data series make them very valuable for nowcasting. We do not use disaggregated data (e.g. sector-level production measures or labor market data broken down in demographic groups) and rely only on the headline indicators for each category.[8] In total, we include 29 time series, which are listed with detailed explanations in Appendix B.3.

2.6 Construction of the real-time database

Macroeconomic data is revised over time by statistical agencies, incorporating additional information that might not be available during the initial releases. In order to mimic the exercise of a real-time forecaster, it is therefore crucial to use vintages of *unrevised* data as they were available at each point in time. We collect real-time vintages spanning the period January 2000 to December 2019 from the Archival Federal Reserve Economic Database (ALFRED). For each vintage, the start of the sample is January 1960, appending missing observations to any series which starts later.

Several intricacies of the data need to be addressed to build a fully real-time data base. For several series, vintages are available only in nominal terms, so we separately obtain the appropriate deflators, which are not subject to revisions, and deflate the series in real time. In several occasions

[6] Additional computational gains can be obtained if the univariate Kalman Filter is compiled in **C**++.

[7] Giannone et al. (2008) conclude that that prices and monetary indicators do not improve GDP nowcasts. Banbura et al. (2012), Forni et al. (2003) and Stock and Watson (2003) find mixed results for financial indicators.

[8] Boivin and Ng (2006), Alvarez et al. (2012), and Banbura et al. (2010, 2012) argue that the strong correlation in the idiosyncratic components of disaggregated series of a given category results in misspecification that can worsen the in-sample fit and out-of-sample forecasting performance of DFMS.

the series are subject to methodological changes and part of their history is deleted by the statistical agency. In this case, we use older vintages to splice the growth rates back to the earliest possible date. For *soft* variables, it is often assumed that these series are unrevised. However while the underlying survey responses are indeed not revised, the seasonal adjustment procedures applied to them do lead to important differences between the series as was available at the time and the latest vintage. We apply the Census-X12 procedure in real time to obtain a real-time seasonally adjusted version of the surveys. We follow the same procedure for initial unemployment claims.

2.7 Real-time forecasting using cloud computing

In the real time dataset, a vintage is constructed for each day in which a new observation or a revision to any of the series is released. On average, this occurs almost 15 times every month. Given that we have 20 years of real-time vintages, this leaves us with approximately 3,600 vintages. We find that our algorithm converges within the first few thousand iterations. We therefore take 7,000 iterations of the Gibbs sampler presented in section 2.4 are run, discarding the first 2,000 as burn-in draws. One such run takes approximately 30 minutes using a high-performance desktop computer.[9] Therefore, the entire exercise across all vintages and different versions of the model would take several months on a single computer. We leverage the possibilities of massively parallelized cloud computation. We have integrated our MATLAB code with the Amazon Elastic Compute Cloud (Amazon EC2), which allows us to compute up to 2,500 runs of the algorithm simultaneously.[10] This reduces the total computation time to just a few hours per model.

3 Implications for modeling and tracking economic activity

This section discusses the implications of our modeling assumptions for the tracking US macroeconomic aggregates as new data arrives. In particular, we highlight the importance of capturing slow-moving trends in long-run growth and output volatility, the evidence for heterogeneity in the responses of macro variables to common shocks, and the implications of fat-tailed observations for

[9] This calculation is carried out using an Intel(R) Core(TM) i7-8700 CPU 3.20GHz with 32.0 GB of RAM.

[10] We are especially grateful to Jago Westmacott and the technology team at Fulcrum Asset Management for developing and implementing a customized interface for seamlessly interacting with the cloud computing platform.

the process of updating beliefs regarding the state of the economy.

3.1 Secular trends: long-run growth and drifting volatilities

Figure 3 displays the estimate of the time-varying long-run growth rate of GDP as well as the SV of the common factor. These estimates condition on the full sample and account for both filtering and parameter uncertainty. Panel (a) presents the posterior median, together with the 68% and 90% high posterior density (HPD) intervals of the long-run growth rate of US real GDP. As discussed in Antolin-Diaz et al. (2017), this estimate conforms with the established narrative about US postwar growth, including the "productivity slowdown" of the 1970's or the 1990s' productivity boom. Importantly, it reveals the gradual slowdown since the start of the 2000's, with most of the decline before the Great Recession. At the end of the sample, there is a slight improvement but the average rate of long-run growth remains just above 2%, highlighting persistently weak activity after the Great Recession (Fernald et al., 2017).

Panel (b) presents posterior estimate of the SV of the common factor. It is evident that volatility declines over the sample, with the Great Moderation clearly visible and still in place, confirming the insights of Gadea-Rivas et al. (2014). In fact, just prior to the COVID-19 pandemic and recession output volatility reached a historically low level of less than 1% in annualized terms. We discuss the implications of the COVID-19 crisis for output volatility in section 5. The plot also shows that volatility spikes during recessions, a feature that brings our estimates close to the findings of Bloom (2014) relating to business-cycle uncertainty. It appears that the random walk specification is flexible enough to capture cyclical changes in volatility as well as permanent phenomena.

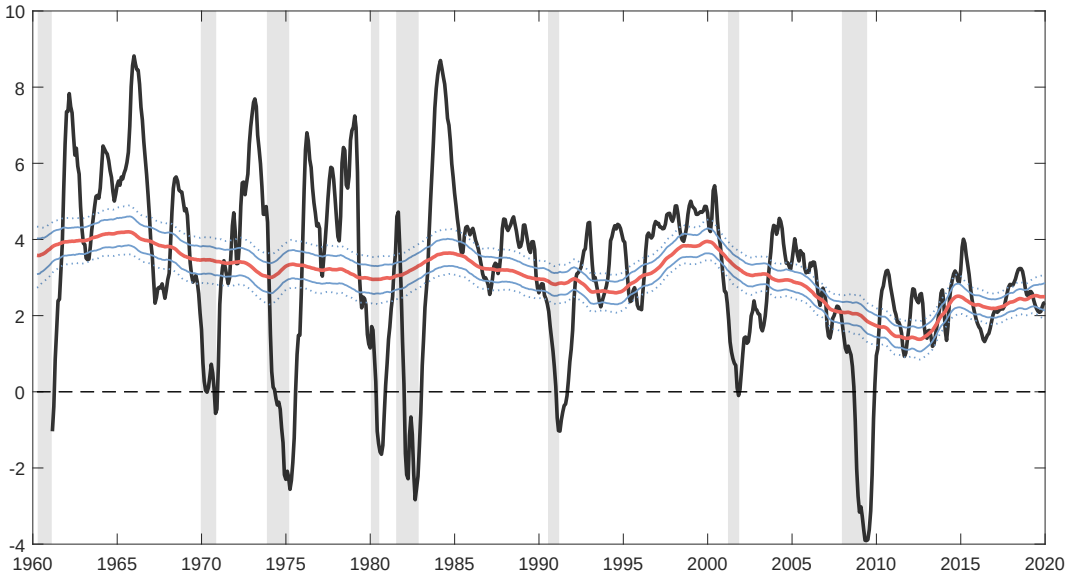
In addition to analyzing the volatility of the common factor, it is possible to compute an index of uncertainty following Jurado et al. (2014) to summarize the uncertainty in each variable. In particular, Figure 4 reports the mean and HPD bands of the uncertainty index

$$\mathcal{U}_t \equiv \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{\lambda_i \sigma^2_{\varepsilon,t}\lambda_i' + \sigma^2_{u,i,t}}{\hat{s}_i^2}} \tag{10}$$

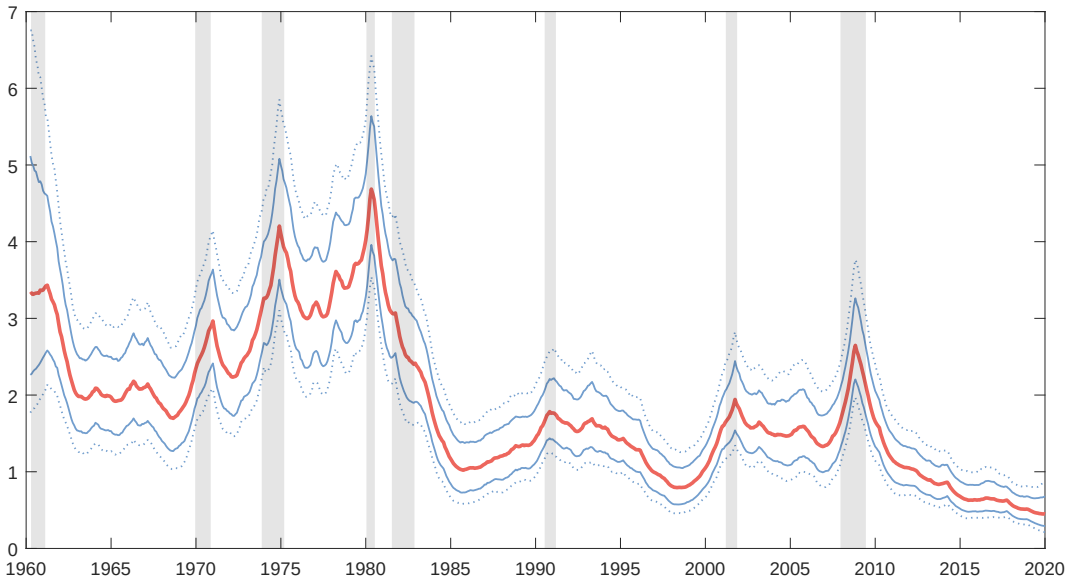
where $\hat{s}_i$ is a variable-specific scaling factor capturing differences in average standard deviation across variables. The advantage of such approach is that, apart from reflecting the time variation

Figure 3: ESTIMATED LONG-RUN COMPONENTS OF US GDP GROWTH

(a) Real GDP growth and estimated long-run growth rate

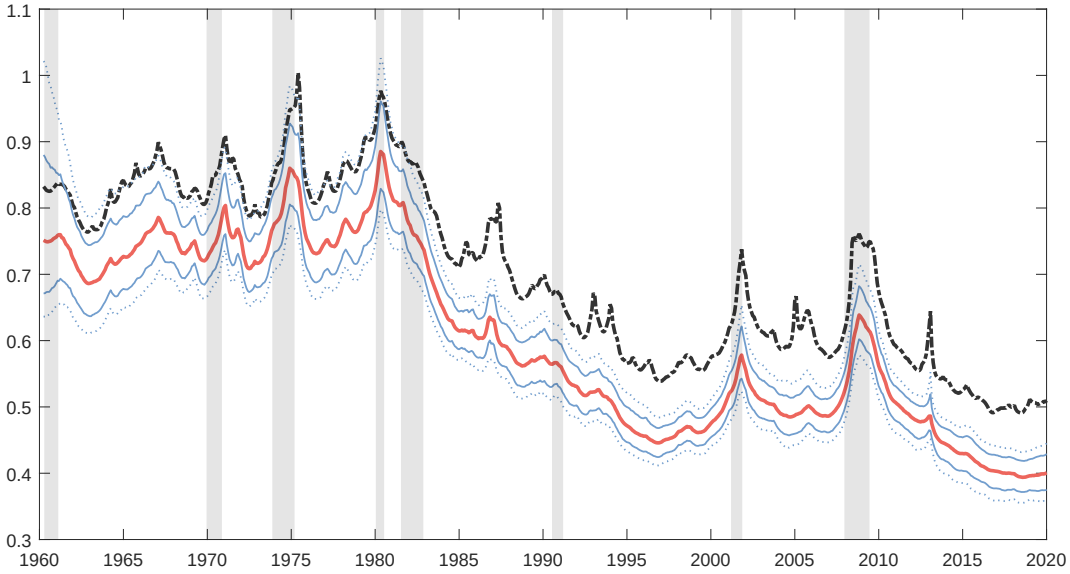


(b) Estimated time-varying volatility of the common activity factor



Notes. Posterior estimates of the low-frequency components explicitly modeled in our Bayesian DFM. Panel (a) plots the growth rate of US real GDP (solid black line) and superimposes the posterior median (solid red) and the 68% and 90% (solid and dotted blue) posterior credible intervals of the time-varying long-run growth rate estimated by the model. The estimate uncovers meaningful time-variation in the long-run growth rate, which lines up familiar episodes of US postwar growth. Panel (b) presents the median (solid red), together with the associated 68% and the 90% (solid and dotted blue) posterior density credible intervals of the volatility of the common factor. This is the square root of $\text{var}(f_t) = \sigma^2_{\varepsilon,t}(1-\phi_2)/[(1+\phi_2)((1-\phi_2)^2-\phi_1^2)]$. Our estimate implies a trend decline in volatility (Great Moderation), as well as heightened volatility in economic contractions. In both panels, gray shaded areas represent NBER recessions. See also Antolin-Diaz, Drechsel, and Petrella (2017) for a detailed discussion related to the patterns visible in both panels.

Figure 4: UNCERTAINTY INDEX



Notes. Posterior mean (solid red) and 68% and 90% (solid and dotted blue) posterior credible intervals of the index of uncertainty following Jurado et al. (2014). This index is computed as shown in equation (10) for the full Bayesian DFM. For comparison, the broken black line in displays the estimate of the same index calculated from a version of the model which does not feature the outlier component. The latter is more similar to the fairly volatile estimate presented by Ludvigson et al. (2015). This suggests caution when interpreting movements in economic uncertainty indices in the presence of variables that feature fat-tailed observations that may not wash out in the aggregate.

in the volatility of the common factor, $\sigma_{\varepsilon,t}$, as in Figure 3 (b), it would capture any unmodeled common component in the volatilities of the idiosyncratic components, $\sigma_{u,i,t}$. Two features are worth noting. First, our estimate displays a marked downward trend associated with the Great Moderation, which contrasts with the one of Jurado et al. (2014) which uses both macro and financial variables, but appears closer to the estimates obtained by Ludvigson et al. (2015) using real activity only. This suggests that the Great Moderation is a phenomenon specific to real economic activity. Second, our estimate displays many less spikes than existing estimates, rarely increasing outside of recessions. This contrasts with the fairly volatile estimates of Ludvigson et al. (2015). The main explanation is our treatment of fat tails. The broken black line in Figure 4 displays the same index calculated from a version of the model which does not feature the outlier component, which exhibits more frequent and larger spikes. Some of them, like the one visible around 2005, can be attributed to short-lived natural disasters such as hurricane Katrina. Our results suggest caution when interpreting movements in economic uncertainty indices when some of the variables

included display fat-tailed innovations which may not wash out in the aggregate.[11]

3.2 Heterogeneous dynamics

The second key departure from the standard factor model is the presence of the lag polynomial $\mathbf{\Lambda}(\mathbf{L})$ of order m in Equation (1). The standard case in which $m = 0$ implies that the impulse response functions (IRFs) to a common shock,

$$\frac{\partial y_{i,t+h}}{\partial \boldsymbol{\varepsilon_t}} = \lambda_i \left(\mathbf{\Phi}(\mathbf{L})\right)^h$$

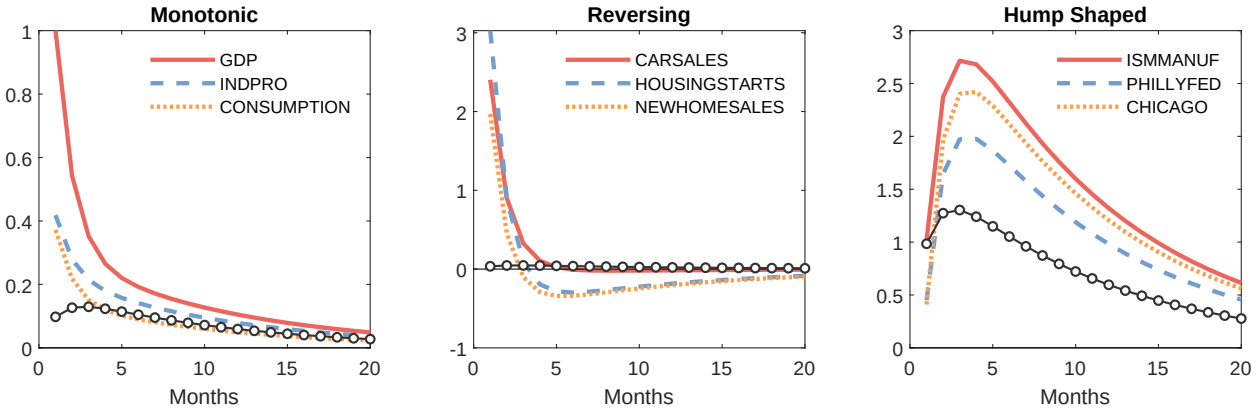
are all proportional across variables, differing only by the factor loadings, λ_i, which acts as a scaling factor. Instead, when $m > 1$ the same common shock can propagate with very different dynamics across variables. D'Agostino et al. (2015) refer to this case as "dynamic heterogeneity" and explore its implications for a six-variable model. The larger size of our dataset allows to uncover three broad patterns of impulse responses, illustrated in Figure 5. First, broad aggregates capturing output and production respond with a decaying pattern, where the peak response is on impact but persists for many months. Second, certain variables, particularly those related to investment such as vehicle sales or construction variables, have a strong initial response which turns negative after a few months before decaying to zero. Recall that the variables are expressed in differences, so this indicates an initial overshoot and subsequent decay in levels. Finally, a number of variables, primarily the business surveys, display clear hump-shaped responses to the common shock.

Heterogeneous IRFs can account for the lead-lag dynamics observed in the data as illustrated in Figure 1 even with contemporaneous aggregate shocks. In a specification that does not include heterogeneous dynamics ($m = 0$), the in-sample cross-correlogram of the common activity factor is lagged relative to the one of GDP, with the peak correlation achieved in the second lag (Figure 6, Panel a), whereas it peaks at lag zero for a model with $m = 1$. The common component of GDP thus better captures contemporaneous movements in activity.

The patterns highlighted above are important for the properties of the DFM and significantly affect its forecasts. The previous literature has typically found that business surveys are estimated to

[11] This is particularly important when the number of variables n, is small or moderate, as is our case, and may be a less important concern when it is very large, as in Ludvigson et al. (2015).

Figure 5: IRFS OF SELECTED VARIABLES TO A COMMON SHOCK



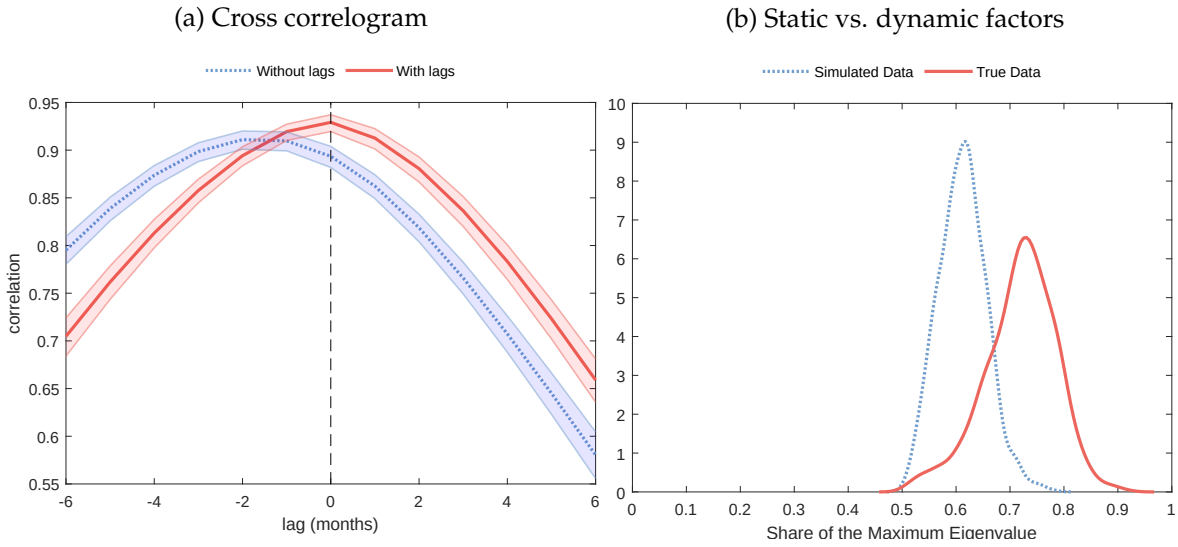
Notes. IRFs of different variables to an innovation in the process of the dynamic factor (that is, an innovation to ε in equation (3)). In each panel, the black line with circular markers shows the IRF of the common factor itself, while the solid red, orange dotted and dashed blue lines show the IRFs for selected variables. The three panels of the figure group these IRFs into categories based on the different shapes of the responses ('monotonic', 'reversing' and 'hump-shaped'), that arise from the possibility of lead-lag responses. In a model with homogeneous dynamics, all of the IRFs would be proportional to the factor by construction, that is, the model would hardwire hump-shaped dynamics into all indicators.

have a high weight in the computation of the factors, due to their high correlation with the business cycle and high degree of persistence. As a consequence, the factor estimated with homogeneous dynamics ($m = 0$) inherits the hump-shaped IRF of these variables. In turn, as the dynamics of the remaining variables are difficult to fit with this functional form, the estimates of their loadings are dampened towards zero, in some cases receiving virtually zero weight in the computation of the factors. This is displayed as the black IRF with circular markers in Figure 5.

An implication of our modelling choice, with a single factor ($k = 1$) and heterogeneous dynamics ($s > 0$) is that we are taking the view that a single aggregate shock is responsible for the bulk of fluctuations in our panel, even though its propagation is different across variables. This parsimonious specification contrasts with an alternative modeling choice, followed, e.g. by Bok et al. (2018), in which more factors are added. The two specifications are related by the fact that a model with heterogeneous dynamics can always be re-written as a model with more than one factor, homogeneous dynamics, and a rank restriction on the variance of the transition equation.[12].

[12] A large part of the literature calls the first specification a "dynamic factor model" and the second a "static factor model" (see, e.g., Stock and Watson, 2005), whereas another part calls the model a "dynamic factor model" whenever the transition equation of the factors contains lags, i.e. $p > 1$. We adhere to the second terminology throughout and refer to the case where $m > 0$ as "dynamic heterogeneity" following D'Agostino et al. (2015).

Figure 6: UNPACKING HETEROGENEOUS DYNAMICS



Notes. Panel (a) present two separate in-sample cross-correlograms between the common activity factor and real GDP growth. One is for a standard DFM which does not allow for lead/lag patterns (blue), the other for our full Bayesian DFM (red). For the former, the peak correlation achieved in the second lag, whereas it peaks at lag zero for the latter. This implies that the common component of GDP better captures contemporaneous movements in activity. Panel (b) is based on rewriting our DFM as a two-factor model (see equation (12)) and estimating it on our data set. We check how close the estimated factor covariance matrix of this model is to reduced rank. This done is to verify whether the restriction on the two-factor model's covariance matrix implied by our heterogeneous dynamic specification is satisfied. In particular, we compute the share of its largest eigenvalue for each posterior draw (solid red curve). The modal share is around 75% of the largest eigenvalue. This is very high, and higher even from an estimate of the same quantity using simulated data where the restriction is satisfied (shown as dotted blue). This analysis supports idea that a single aggregate shock, transmitted heterogeneously across variables, drives the common fluctuations in our data panel.

As an illustrative example, consider a model with a single factor ($k = 1$) and where the factor and its first lag load on the variables ($m = 1$). Its transition equation is written

$$\begin{bmatrix} f_t \\ f_{t-1} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_{t-1} \\ f_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_t, \tag{11}$$

A single shock η_t is responsible for the common fluctuations in the observed variables. Instead, the transition equation with two factors and homogeneous dynamics would be written

$$\begin{bmatrix} F_t^1 \\ F_t^2 \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} F_{t-1}^1 \\ F_{t-1}^2 \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{12}$$

The key insight is that (12) can be rotated into (11) with the additional restriction that the covariance matrix of ϵ_t has reduced rank. Which specification is preferred amounts to asking how close to reduced rank is the covariance matrix of specification (12) in the data. We estimate specification (12) with an unrestricted covariance matrix for our data set, and compute, for each posterior draw, the share of the largest eigenvalue in the factor covariance matrix. The result, displayed in panel (b) of Figure 6, is a modal share of about 75% of the largest eigenvalue. This is very high, and higher even from an estimate of the same quantity using simulated data where the restriction is satisfied. Our results indicate strong support for the idea that a single aggregate shock, transmitted heterogeneously across variables, drives the common fluctuations in our panel.

3.3 Interpreting the data flow in the presence of fat tails and outliers

Finally, we discuss the impact of fat tails and outliers for the process of updating beliefs about the current state of economic activity as new data arrives, a process which Banbura et al. (2010) refer to as "following the real-time data flow". The events of the first two decades of the twenty-first century have made clear that the Gaussian assumption typically employed in DFMs may not be adequate for the modeling of modern macroeconomic fluctuations, where large shocks and changes in volatility are recurrent.

In principle, there are two sources of fat tails in macroeconomic data. First, *aggregate* increases in volatility for instance caused by a large recession, which leads to outsize movements in many variables. This departure from Gaussianity is captured by the stochastic volatility in the innovations to the common factor (7), which implies an unconditionally fat-tailed distribution of the common factor. The second source of fat tails are recurrent, very large, idiosyncratic movements in specific series, for example caused by a strike or a weather-related disruption in a specific sector of the economy. These movements are typically short-lived, with the series returning to their previous levels within one month. Our model captures this second source of fat tails with the outlier component of equation 1. As we shall see in our application to the COVID-19 crisis, the events surrounding the pandemic and subsequent lockdown are a combination of both phenomena.

The implications of fat tails for nowcasting can be framed in terms of the "influence function" in the robust statistics literature which describes the effect of an additional observation on a statistic of

interest Hampel et al. (1986).[13] In a linear Gaussian model the influence function is linear and when the statistic of interest is the factor estimate it measures the revision of the latter after the arrival of news and is therefore often labelled "news decomposition" (Banbura and Modugno, 2014):

$$E(f_{t_k}|\Omega_2) - E(f_{t_k}|\Omega_1) = \mathbf{w_j}\left(y_{j,t_j} - E(y_{j,t_j}|\Omega_1)\right) \tag{13}$$

where Ω_2 is an information set containing additional observations relative to information set Ω_1, and $y_{j,tj} - E(y_{j,tj}|\Omega_1)$ is the *news*, the forecast error using the old information set.[14] The weight $\mathbf{w_j}$ is the slope of the influence function, and can be shown to have the form

$$\mathbf{w_j} = \frac{E\left((f_{t_k} - f_{t_k|\Omega})(f_{t_j} - f_{t_j\Omega})\right)\Lambda_j'}{\Lambda_j E\left((f_{t_j} - f_{t_j|\Omega})(f_{t_j} - f_{t_j\Omega})\right)\Lambda_j' + \sigma^2_{\eta_{j,tj}}} \tag{14}$$

The expectation term in the denominator is the uncertainty about the common factor, which increases with the stochastic volatility of factor innovations. The derivative of the weight with respect to this term is positive, implying that news lead to larger revisions in beliefs about the state of economic activity in periods of high volatility. Therefore, stochastic volatility in the factor induces time variation in the influence functions. It follows that an increase in the presence of large errors across many variables will lead to an increase in the estimated time-varying volatility of the common factor, and an increased sensitivity of the factor estimates to additional news.

In turn, the Student-t distributed outliers modify the influence function so that the weights are no longer linear, but depend on the value of the forecast error itself:

$$\mathbf{w_j}(y_{j,tj}) = \frac{\Lambda_k E\left((f_{t_k} - f_{t_k|\Omega})(f_{t_j} - f_{t_j\Omega})\right)\Lambda_j'}{\Lambda_j E\left((f_{t_j} - f_{t_j|\Omega})(f_{t_j} - f_{t_j\Omega})\right)\Lambda_j' + \sigma^2_{\eta_{j,tj}}\delta_{j,tj}} \tag{15}$$

where

$$\delta_{j,tj} = (((y_{j,tj} - E(y_{j,tj}|\Omega))^2/\sigma^2_{\eta_{j,tj}} + v_{o,j})/(v_{o,j} + 1) \tag{16}$$

[13] Loosely speaking, the influence function is the first derivative of a statistic T at an underlying distribution F, where the point x plays the role of the coordinate in the infinite-dimensional space of probability distributions (see Hampel et al., 1986).

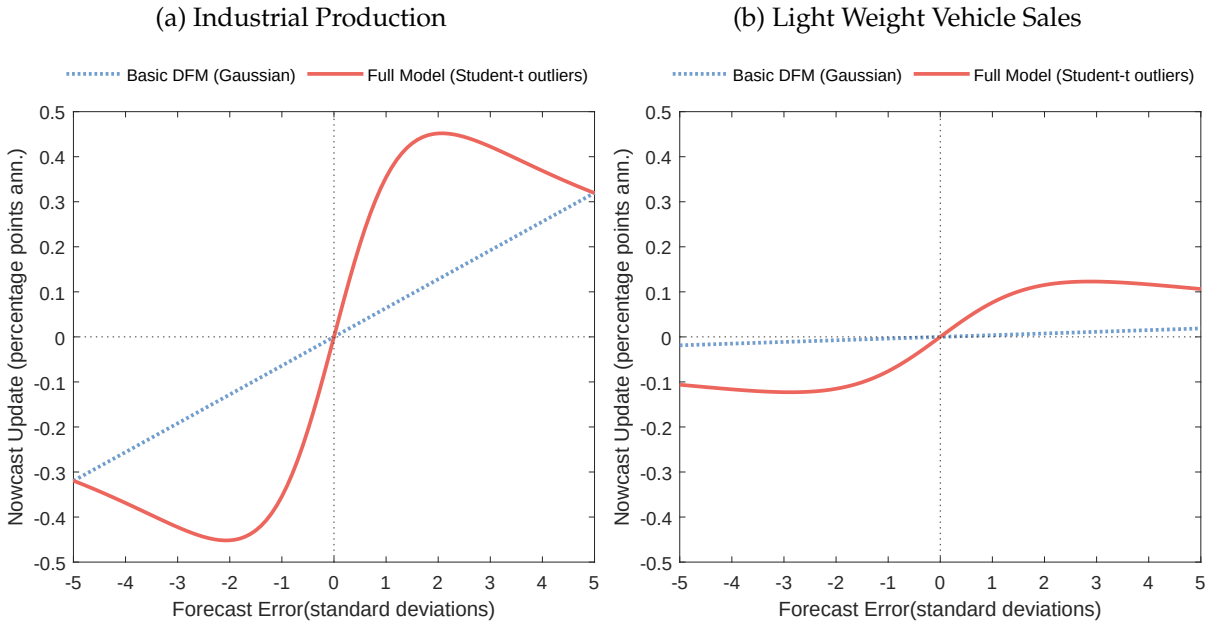
[14] The notation t_k and t_j allow for the time periods for which the news about variable j refers to, and the relevant observation for the factor, t_k to differ.

where $v_{o,j}$ is the estimate of the degrees of freedom of the t distribution of the outlier of variable j. Therefore, large idiosyncratic errors are discounted as outlier observations containing less information. Figure 7 plots the influence functions for an example variable (industrial production) in the linear case (without outliers) and the nonlinear case (with outliers). For a model with Gaussian innovations, the update of the factor is a line with slope equal to $w_{i,j}$. These are displayed as the dotted blue lines in Figure 7.[15] As can be seen, noisy variables such as car sales have very low weights, meaning that surprises to this variable lead to small updates to the current factor. With t-distributed outliers, the influence functions are now the S-shaped red lines of Figure 7. Around the origin, i.e., for a small surprise, the function is close to linear, but as the surprise increases in size the update of the factor tapers off, and eventually can decrease in size. The intuition is clear: if one observes a three- or four standard deviation surprise, it is increasingly likely that that observation represents a one-off outlier in the data, and therefore our estimate of underlying economic activity should respond less to those "news". Nevertheless, the update is not zero, which would be the case if the outlier would be replaced with a missing observation.

It is also worth noting that the heterogeneous lead-lag dynamics outliers interact with each other. Allowing for heterogeneous dynamics increases the relative weight of "hard" variables, such as industrial production and car sales, relative to business surveys (this implies an increase in the slope of the influence function). However, these series are precisely the ones in the panel that are likely to feature outliers. A model with heterogeneous dynamics but no outlier component would thus produce highly volatile revisions to the estimated factor in response to transitory movements in hard data. The combination of both features allows a model which gives weight to the hard data for small surprises but is not unduly influenced by transitory outliers.

[15] The influence functions for all variables are available in the Appendix.

Figure 7: INFLUENCE FUNCTIONS (BELIEF REVISION AS A FUNCTION OF FORECAST ERROR)



Notes. The panels of the figure plot the *influence functions* for two variables, industrial production and light weight vehicle sales. This function indicates by how much the estimate of the dynamic factor is updated when the release in the variable is different from its forecast and thus contains "news". The dotted blue line plot these influence functions in the Gaussian case, as in equation (14), while the red lines represent the student-t case. As shown in equations (15)-(16) the model with fat tails allows these functions to be nonlinear and nonmonotonic. It is also noteworthy that in the case of vehicle sales, only the full DFM assigns any importance to the release for updating the factor. As explained in the text, this is due to the combination of lead-lag dynamics and the Student-t component. The influence functions for other variables are provided in the Appendix.

4 Real-time out-of-sample model evaluation

This section presents the out-of-sample evaluation of the model. As described in Sections 2.6 and 2.7, this evaluation is carried out on a fully real-time data base which includes unrevised vintages of the data series for each point in time. Our desire is to mimic the exercise of a forecaster updates her information set and produces a nowcast of GDP in real-time every day from January 2000 to December 2019. Our evaluation results are presented as follows. In Section 4.1, we compare the GDP nowcasting performance of the full model to a benchmark DFM which does not feature our three methodological innovations. We consider point forecasts but also pay special attention to the production of well-calibrated density forecasts that accurately characterize the uncertainty around current economic activity. In Section 4.2, we decompose the performance of the model's forecasting

performance into each of the new features individually. This provides a comprehensive assessment of the importance of the individual contributions we make to the nowcasting process.

4.1 Evaluation results: point and density forecast accuracy

We evaluate the point and density forecast accuracy relative to the third release. As GDP gets revised over time, an important question is which vintage of GDP is taken as the "ground truth" against forecasts are evaluated. In particular, the Bureau of Economic Analysis releases an "advance" estimate between 25 and 30 days after the end of the reference quarter, a "second" estimate one month after that, and a "final" one yet another month after the second estimate. After that, GDP estimates continue being revised for many years as annual revisions and other improvements are incorporated over time. We focus our evaluation on the third (or "final") release, as the majority of revisions occur in the first three releases.[16]

We test the significance of any improvement of our full model relative to a DFM that is estimated on the same data set but does not feature time-varying trends and SV, heterogeneous dynamics or fat tails, such as the model of Banbura et al. (2010).[17] Figure 8 shows the accuracy of the two models in predicting the third release of GDP. We evaluate this accuracy starting 180 days before the end of the reference quarter (a one quarter ahead prediction), as the reference quarter unfolds (a nowcast) and up to 30 days after the end of the quarter (a backcast), at which point the advance release is usually published. Panel (a) presents the root mean squared error (RMSE) as a measure of point forecasting accuracy, whereas Panel (b) evaluates the density forecasting accuracy using the Log Score. An evaluation based on alternative measures – mean absolute error (MAE) and continuous rank probability score (CRPS) – can be found in Appendix C.

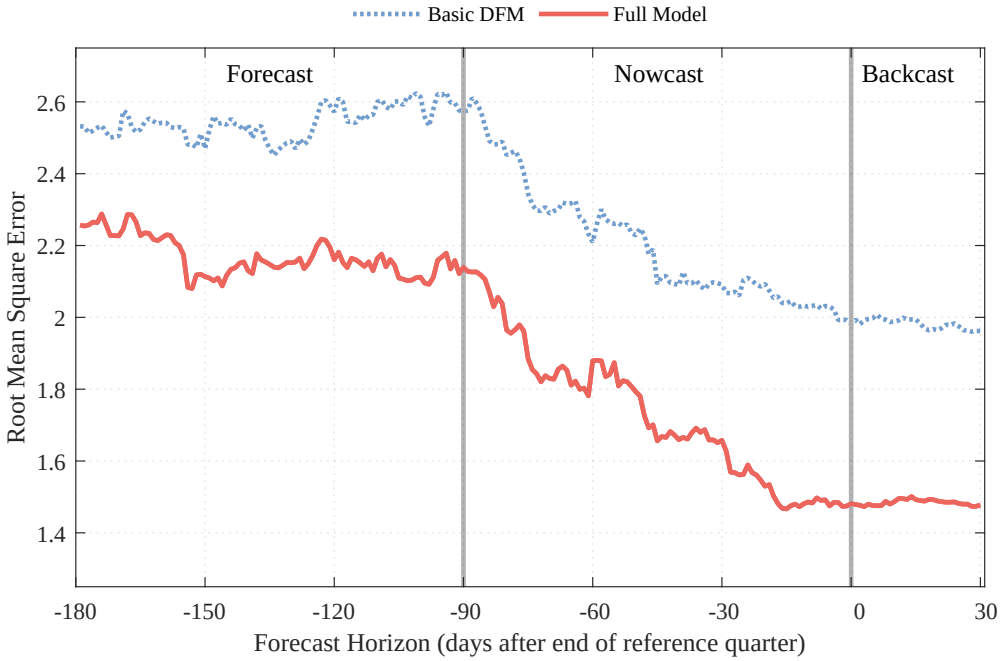
Panel (a) shows that RMSE of both models declines as forecast horizon gets shorter and information contained in monthly indicators becomes available. Importantly, the RMSE of the

[16] We have explored the alternative of evaluating the forecasts against earlier or subsequent releases, the latest available vintages, or an average of the expenditure and income estimates of GDP. The relative performance of the models is broadly unchanged, but we find that all models do generally better at forecasting less "mature" vintages. If the objective is to improve the performance of the model relative to the first official release, then ideally an explicit model of the revision process would be desirable. The results are available upon request.

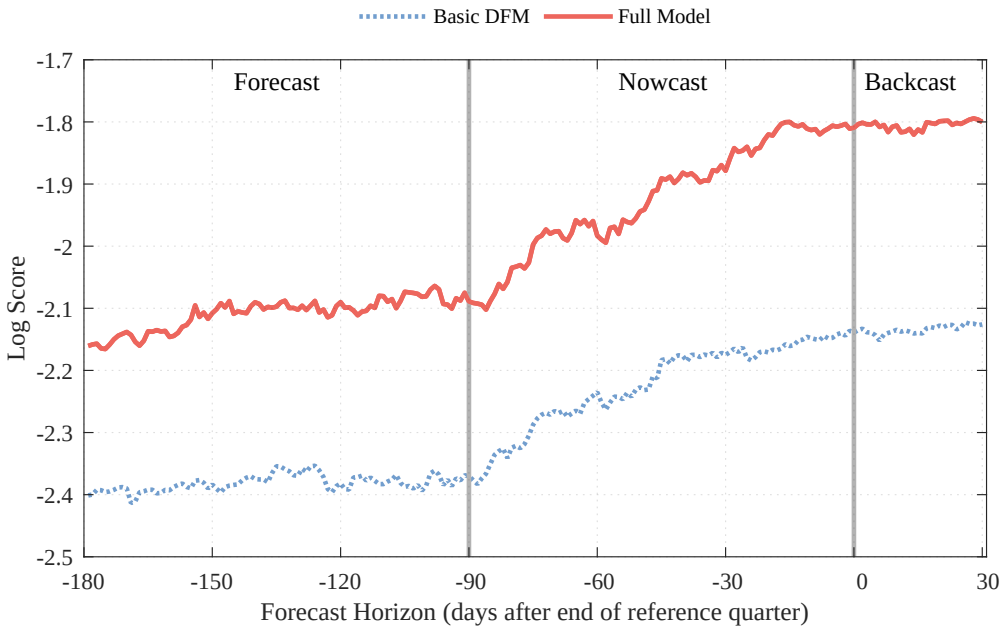
[17] We follow the "pragmatic approach" to testing statistical significance of Faust and Wright (2013) and Groen et al. (2013). These authors build on Monte Carlo results in Clark and McCracken (2012) and show the applicability of the Harvey et al. (1997) small-sample correction for the Diebold and Mariano (1995) statistic to nested and non-nested models.

Figure 8: GDP FORECAST EVALUATION AS THE DATA ARRIVES

(a) Point forecast accuracy: RMSE (accurate forecast implies low value)



(b) Density forecast accuracy: Log score (accurate forecast implies high value)



Notes. In both panels, the horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter of a given GDP release. Thus, from the point of view of the GDP forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of the next quarter; forecasts at a 90-0 day horizon are nowcasts of the current quarter, and the forecasts produced 0-25 days after the end of the quarter are backcasts of the last quarter. Panel (a) plots the RMSE of the full model (solid red line) as well as the basic DFM (dotted blue) over this horizon. A more accurate forecasts implies a lower RMSE. Panel (b) displays the analogous evolution of the log score of the two models, a measure of density forecast accuracy, which takes a higher value for a more accurate forecast. The differences between the models are statistically significant throughout the horizon in both panels. See also Table 1.

full model is much lower than that of the basic model throughout the horizon, and in particular the model delivers much more accurate forecasts over the nowcasting period. The Diebold and Mariano (1995) test indicates that the difference is statistically significant at all horizons at the 5% level and becomes significant even at the 1% level from around 120 days before the end of the reference quarter. As highlighted by Banbura et al. (2012), an important property of an efficient nowcast is that the forecast error declines monotonically over time as new information arrives. We observe that for the basic and the full model, the majority of the valuable information comes within the nowcast quarter, with the accuracy improvements stabilizing around the end of the reference quarter (horizon zero), but the decline in RMSE is steeper for the full model, implying a more efficient use of incoming information. In summary, the added features result in reducing the RMSE by around half a percent.

Panel (b) turns to an evaluation of density forecasts. Density forecasts evaluate the accuracy of the entire predictive distribution, instead of focusing exclusively on its center. They are used to predict unusual developments or tail risks, such as the probability of a recession or a strong recovery given current information. Our Bayesian framework allows us to produce such density forecasts, consistently incorporating filtering and estimation uncertainty. There are several measures available for the formal evaluation of density forecasts. We focus on the (average) log score, which is the the logarithm of the predictive density evaluated at the realization. This rewards the model that assigns the highest probability to the realized events and is a popular evaluation metric.[18] Panel (b) shows that our model outperforms its counterpart also in terms of density forecasting at all horizons. The Diebold and Mariano (1995) test indicates that the difference in performance is significant at the 1% level at all horizons.

The results presented so far document the average performance over the period 2000-2019. It is useful to examine whether the out-performance of the full is stable over time or due to a few special periods. Figure 9 presents, for a fixed forecast horizon, the relevant loss function computed recursively over time. We chose the middle of the nowcasting quarter (45 days before the end of the reference period), but choosing a different horizon would tell the same story: for both point and density forecast, the improvement in performance is stable across time and would have been

[18] We report results based on the (average) continuous rank probability score (CRPS) in Appendix C.

clear just a few years after the start of the evaluation period. It is also the case that some of the out-performance of the full model appears to happen in the period just after recessions, suggesting the full model is more accurate at capturing the dynamics of recoveries, a point to which we return in the next section.

Figure 10 compares the predictions of the two versions of the model with the actual release of GDP. In this case we fix the forecast horizon to the day before the advance release is published.Both models track the broad contour of fluctuations in real GDP, including the two recessions and recoveries in 2001 and 2008-09. However it is clear that, particularly after 2010, the basic model displays a persistent upward bias, reflecting a failure to capture the decline in long-run growth experienced in this decade, as documented by Antolin-Diaz et al. (2017). It is also visible that the full model correctly anticipated stronger recoveries to both recessions.

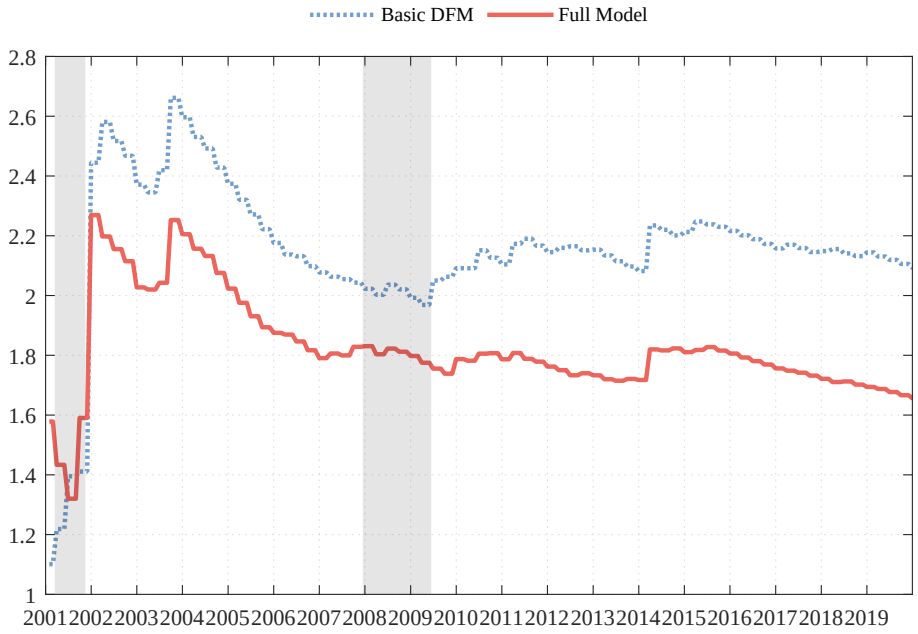
Turning to density forecasts, the shaded areas in the figure represent a 68% HPD interval for each model. Thus, if the density forecasts were well calibrated, out of the 80 quarterly observations we would expect about 26 of them, or 32%, to fall outside of the bands. For the case of the basic model, 17 or 21% fall outside, whereas for the full model it is 27, or 34%. This indicates that the basic model, which does not feature time-varying volatility, is on average too conservative, producing bands that are too wide. The restrictive assumption of constant volatility is behind this result: the sample features long, stable expansions punctuated by relatively large recessions, so a single, average, estimate of volatility is an overestimation most of the time.

4.2 Decomposing the contribution of the novel components

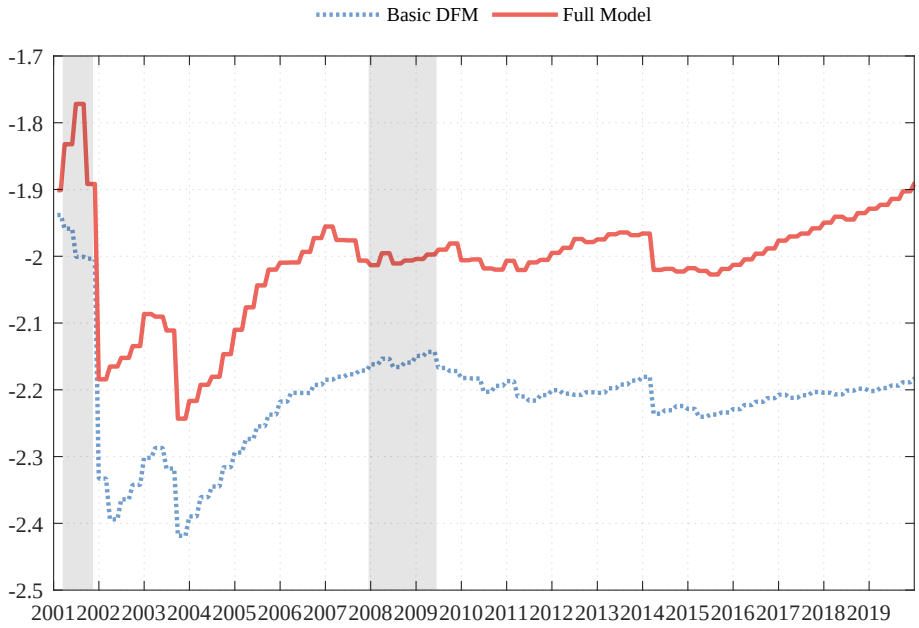
We now examine which of our modelling innovations are responsible for the overall improvement in performance. Table 1 reports the evaluation exercise for a few different models of increasing complexity: first, for comparison, a simple autoregressive (AR) model of order 1 estimated on quarterly data; second, the basic DFM described above; third a version which adds only time-varying long-run growth and stochastic volatility (as in Antolin-Diaz et al., 2017), labeled "trend & SV"; fourth, a version which adds, on top, the heterogeneous dynamics ("Lead/Lag"), and finally the full model which adds on top the t-distributed outliers ("fat tails"). We report the performance in RMSE (top panel) and Log score (bottom panel) for various forecasting horizons.

Figure 9: GDP FORECAST EVALUATION THROUGH TIME

(a) Point forecast accuracy: RMSE (accurate forecast implies low value)

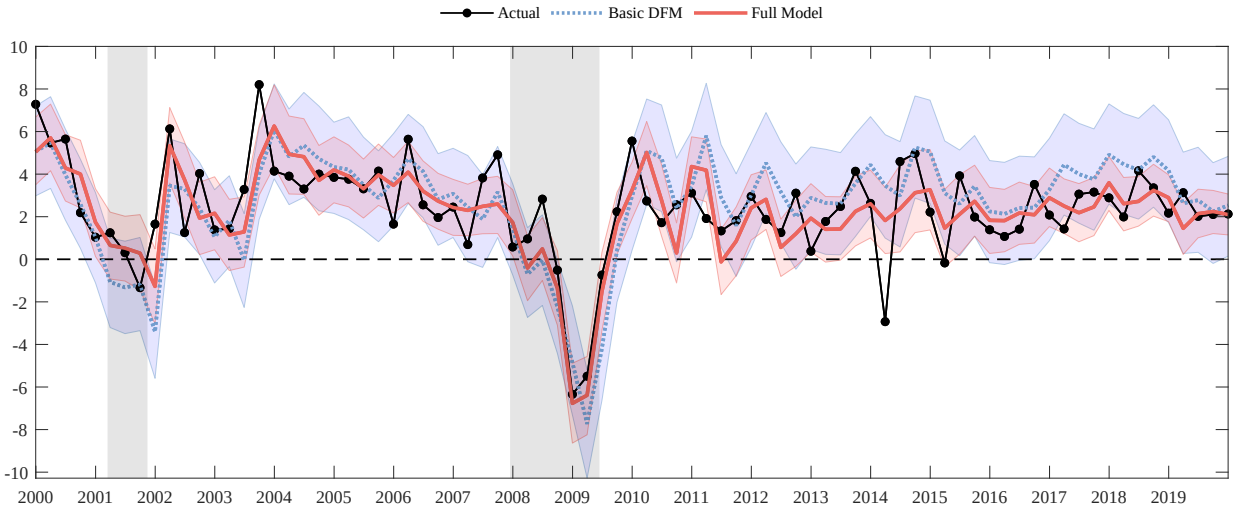


(b) Density forecast accuracy: Log score (accurate forecast implies high value)



Notes. In both panels, the horizontal axis indicates the time of the evaluation sample 2000-2019. Panel (a) plots the rolling RMSE of the full model (solid red line) as well as the basic DFM (dotted blue) through. A more accurate forecast implies a lower RMSE. Panel (b) displays the analogous rolling evolution of the log score of the two models, a measure of density forecast accuracy, which takes a higher value for a more accurate forecast. In both panels, the gray shaded areas indicate NBER recessions. The rolling metrics indicate that the full DFM is preferred based on both point and density forecasting performance as early as 2002.

Figure 10: MODEL NOWCASTS AND PUBLISHED GDP GROWTH COMPARED



Notes. The black line represents the time series of actual real GDP growth in the United States. The blue line plots the nowcasts for real GDP growth based on information up to this point our full model (red) and the basic DFM (blue). The blue and red shaded areas indicate the corresponding density around the nowcasts. It is visible that the nowcasts for the basic model exhibit an upward bias in the later part of the sample, which is not the case in the full model. The full model also appears to be better at capturing turning points in recessions.

Starting with point forecasting, and looking from left to right, we see how the performance improves across model versions. The basic DFM struggles to improve on the simple AR(1) in terms of RMSE, but the addition of the time-varying long-run trend and volatility leads to a reduction in RMSE at all horizons.[19] The most substantial improvement comes from the addition of heterogeneous dynamics, whereas the fat tails, unsurprisingly, contribute little to point forecasting accuracy. Turning to the lower panel, which reports density forecast accuracy, we see how the DFMs are substantially better than the AR(1), in particular once the long-run growth and stochastic volatility are added, at which stage the improvement is large and highly significant. The introduction of heterogeneous dynamics leads to an additional large improvement.

In sum, the results of the out-of-sample evaluation exercise indicate that, using data available in real time for the period 2000-2009, the full DFM substantially outperforms its basic counterpart (as well as an even simpler AR(1) model) at all horizons. The improvement in density forecasting is particularly notable, and the trend, the stochastic volatility, and the heterogeneous dynamics

[19] This conclusion is somewhat unfair for the basic DFM, as the AR(1) model is relatively accurate during expansion periods, but fails to track GDP in periods of recession, where the forecasts are arguably more valuable.

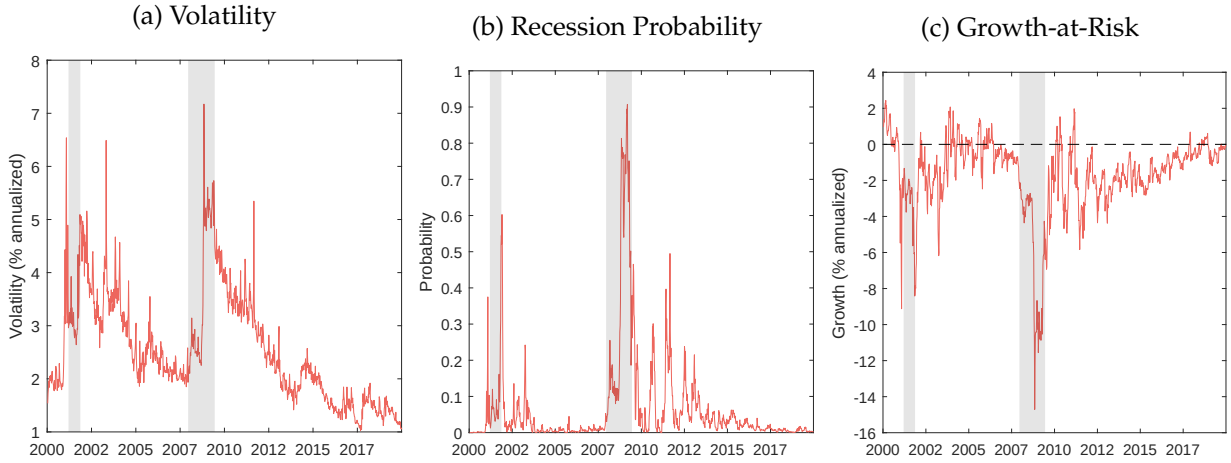
Table 1: FORECASTING PERFORMANCE OF INTERMEDIATE MODEL VERSIONS

RMSE across horizons [p-value against AR(1)]	**AR(1)**	**Basic DFM**	**Trend & SV**	**Lead/lag**	**Fat tails**
-180 days	2.4	2.5	2.4	2.3	2.3
		[0.80]	[0.63]	[0.37]	[0.36]
-90 days (start reference quarter)	2.3	2.6	2.4	2.2	2.1
		[0.75]	[0.59]	[0.30]	[0.28]
-60 days	2.1	2.2	2.0	1.9	1.9
		[0.58]	[0.39]	[0.25]	[0.23]
-30 days	2.1	2.1	1.9	1.7	1.7
		[0.48]	[0.29]	[0.12]	[0.10]
0 days (end reference quarter)	2.1	2.0	1.8	1.5	1.5
		[0.33]	[0.18]	[0.05]	[0.04]
+30 days (first release)	2.1	1.9	1.8	1.5	1.5
		[0.28]	[0.13]	[0.04]	[0.03]
Log score across horizons [p-value against AR(1)]	**AR(1)**	**Basic DFM**	**Trend & SV**	**Lead/lag**	**Fat tails**
-180 days	-2.42	-2.40	-2.16	-2.16	-2.16
		[0.36]	[0.00]	[0.00]	[0.00]
-90 days (start reference quarter)	-2.40	-2.37	-2.21	-2.10	-2.09
		[0.33]	[0.03]	[0.00]	[0.00]
-60 days	-2.34	-2.24	-2.05	-1.99	-1.98
		[0.09]	[0.00]	[0.00]	[0.00]
-30 days	-2.34	-2.18	-2.00	-1.90	-1.88
		[0.02]	[0.00]	[0.00]	[0.00]
0 days (end reference quarter)	-2.34	-2.14	-1.98	-1.83	-1.81
		[0.00]	[0.00]	[0.00]	[0.00]
+30 days (first release)	-2.35	-2.13	-1.96	-1.81	-1.80
		[0.00]	[0.00]	[0.00]	[0.00]

Notes. Comparison of the forecasting performance across model versions with increasing complexity: a simple AR(1) model estimated on quarterly data; the basic DFM described above; a version which adds only time-varying long-run growth and stochastic volatility (as in Antolin-Diaz et al., 2017), labeled "trend & SV"; a version which adds, heterogeneous dynamics ("Lead/Lag"); and our full Bayesian DFM which adds on top the t-distributed outliers ("fat tails"). We report the performance in RMSE (top panel) and Log score (bottom panel) for various forecasting horizons. The p-value of a formal test against the model of one lower level of complexity are reported in brackets.

all contribute to the improvements. The fat appear to lead only to a small improvement in the 2000-2019 sample, for which GDP did not feature substantial outliers, but as we will see in the next section will play an important role during the COVID-19 crisis.

Figure 11: REAL-TIME RISK ASSESSMENT MEASURES



Notes. Panel (a) displays a real-time measure of uncertainty, defined as the difference between the 16th and the 84th percentiles of the daily activity measure in Figure 2. Panel (b) plots the probability of recession, defined as the model-implied probability that the current and next quarter GDP growth will be negative. Panel (c) displays at a measure of GDP growth at risk in the spirit of Adrian et al. (2019), measuring the mean below the 5% distribution of the GDP distribution for the current quarter.

4.3 Real-time assessment of economic activity, uncertainty and tail risks

Finally, we derive from the model a daily measure of real economic activity, as well as corresponding uncertainty and tail risk measures. We construct a daily measure of activity by taking a weighted average of the model's estimate of quarterly GDP growth, where the weights vary depending on the day of the month. In this way, we create a rolling measure of current-quarter economic activity that can be updated every day.[20] This is the quantity displayed, together with a 68% HPD credible interval, in Figure 2. The figure also displays the daily real-time estimate of the long-run growth rate of GDP. As can be seen from the figure, the daily indicator tracks the major business cycle events in the last two decades, including the early 2000s recession, the Great Recession of 2008-2009 and the subsequent weak recovery associated with a decline in long-run growth. The probabilistic nature of our model allows us to compute additional statistics related to uncertainty and risk around our

[20]More specifically, recall that for each day τ in the evaluation sample (from January 11 2000 to December 31st 2019) the model is re-estimated using the vintage of information available up to that day, denoted Ω_τ. From equation (1), define the underlying monthly growth rate of GDP as $GDP_t^\star \equiv c_{1,t} + \lambda_1(L)f_t$, i.e. GDP excluding the idiosyncratic and outlier components. Applying to this the Mariano-Murasawa polynomial in (9) we obtain a version of this series expressed as a quarterly growth rate, $GDP_t^{\star,q}$. Then, our daily indicator of real economic activity is a weighted average of the current and next two month's values for underlying quarterly GDP, where the weights are the proportional to the number of days separating τ from the end of the quarter.

daily estimate of economic activity. In particular, Figure 11 plots three real-time measures of risk. Panel (a) displays a real-time measure of uncertainty, defined as the difference between the 16th and the 84th percentiles of the activity measure in Figure 2. This is related to the volatility estimate displayed in 3, but computed in real time. Therefore, it can be interpreted as a measure of business cycle uncertainty as perceived at each point in time by an observer with access to our model. Panel (b) shows at the probability of recession, defined as the model-implied probability that the current and next quarter GDP growth will be negative. Finally, panel (c) presents a measure of GDP growth at risk in the spirit of Adrian et al. (2019), measuring the mean below the 5% distribution of the GDP distribution for the current quarter. All these measures are useful characterizations of the risks around economic activity that go beyond the information contained in central estimates. The good performance in density forecast exhibited by the full model, as shown in the previous subsections, gives us confidence that they can be relied upon in practice.

5 An application to the COVID-19 crisis

Shortly after the first versions of this paper were circulated, the COVID-19 pandemic and associated recession hit the United States. This episode is unprecedented in many ways: first, the *scale* of the drop in economic activity in response to the public health interventions intended to contain the virus, unseen in postwar history; second, the *speed* at which events unfolded, which poses challenges for models that rely only on the usual monthly economic indicators; and third the unique sectoral *composition* of the shock, which unlike previous recessions has hit sectors such as services which usually display little business cycle fluctuations.[21] We now put our model to the test with 2020 data. We emphasize that none of the priors, settings, or modeling choices were altered ex-post to better fit the 2020 data, which means the exercise is purely an out-of-sample test. Our first observation is that some of the data points produced during this period are so extreme, that attempts to estimate the basic DFM, without t-distributed outliers lead to unstable estimates and, often, failure of the MCMC algorithm to converge. Instead, the full model appears to deliver stable

[21] The challenges faced by existing econometric models estimated on 2020 data have been highlighted by a nascent literature which includes Primiceri and Tambalotti (2020); Lenza and Primiceri (2020), Schorfheide et al. (2020), and Cimadomo et al. (2020).

estimates throughout the period.[22] Our second observation is that the *timing* of the shock, which mostly occurred in mid-March, is particularly challenging for models that use monthly data only, as the data release calendar is such that the most timely indicators are released around the turn of each month, referring to conditions prevailing in the previous month. Therefore, it would not be fully captured until indicators for March and April are released at the end of April. As a consequence, a number of "alternative" high-frequency indicators have appeared that offer a partial but more timely reading of economic activity. We next discuss how these can be incorporated into the DFM framework.

5.1 Incorporating higher-frequency data sources

In response to the pandemic a number of novel indicators have been made public that have the potential to give a more timely reading of economic conditions. While Lewis et al. (2020) focus on creating an index of weekly indicators using series that have now been available for more than a decade, we propose a method for incorporating novel series which are available at weekly or daily frequency but have an extremely short history, sometimes starting as late as March 2020, into the DFM framework. The problem of incorporating these series into the DFM framework can be broken into two: first, how to go from the daily or weekly frequency of the novel series to the monthly frequency of the DFM; second, how estimate the model with series of extremely short history. For the first, we follow the simple approach advocated by Knotek and Zaman (2019), and average the available observations for the month at each point in time. This leaves us with monthly time series that are sometimes as short as one or two observations, so estimation of the parameters in equation (1) is clearly infeasible. Our proposal is instead to match these novel series with traditional monthly indicators that are conceptually very close to the former, and condition the most recent value of the traditional indicator with the information in the novel one. From a Bayesian perspective, this can be thought of imputing the fitted value from a regression with a dogmatic prior of zero intercept and unit slope.[23] Of course, the true relationship between the

[22] To enable the basic DFM to run, we need to adopt an ad-hoc procedure to censor outliers used, e.g. by Stock and Watson (2012), discarding observations more than 10 times larger in absolute value from the interquartile range of the series. This implies throwing away a large fraction of the data related to this period.

[23] A disadvantage of our approach is that we are shutting down the uncertainty about the observation which we are imputing. As discussed by however, Antolin-Diaz et al. (2020), conditional forecasting is equivalent to entropic tilting

Table 2: HIGH-FREQUENCY PROXIES AND MATCHED TRADITIONAL SERIES

Monthly Indicator	Start Date	High Frequency Proxy	Freq.	Start Date	Estimated
Real Consumption (excl. durables)	Jan 67	Credit Card Spending (OI)	D	Jan 20	N
Payroll Empl. (Establishment Survey)	Jan 47	Homebase	D	Mar 20	N
Civilian Empl. (Household Survey)	Feb 48	Dallas Fed RPS	BW	Apr 20	N
Unemployed	Feb 48	Dallas Fed RPS	BW	Apr 20	N
Initial Claims for Unempl. Insurance	Feb 48	Weekly Claims (BLS)	W	Jan 67	N
U. of Michigan: Consumer Sentiment	May 60	Rasmussen Survey	D	Oct 04	Y
Conf. Board: Consumer Confidence	Feb 68	Rasmussen Survey	D	Oct 04	Y
U.S. Vehicle Miles Traveled	Jan 70	Apple Mobility Trends	D	Jan 20	N

Notes. List of high-frequency series and related traditional macroeconomic indicators. The high-frequency series are data sources which have received close attention during the COVID-19 induced recession. They have the drawback that they span only a short history. We identify related traditional indicators, which capture similar economic concepts, but are available for a longer time period and therefore get a more meaningful weight in the DFM.

indicators may not be one-to-one, so the relevant empirical question is whether the extra timeliness added by the high-frequency information, and the fact that the DFM is averaging across a number of series, offsets the bias introduced by this approach.

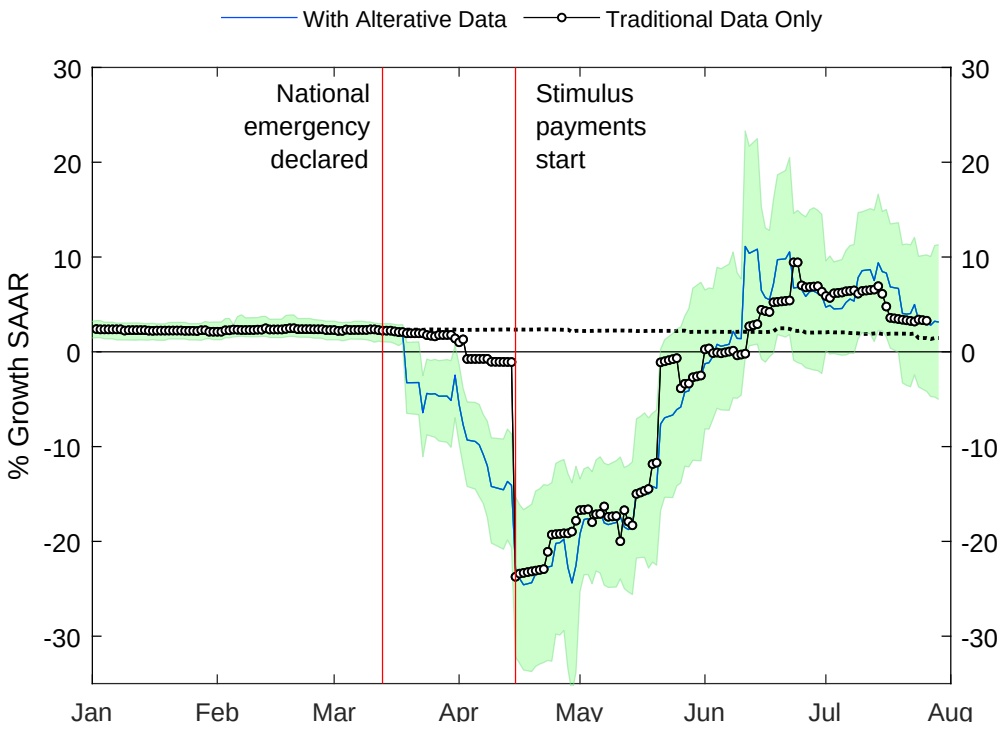
Table 2 lists the novel data series together with the traditional series that we match to each of them. We use the credit card spending series from Chetty et al. (2020) to condition the value of real consumption growth from the BLS; employment data from Homebase and the Real Time Population Survey (Bick and Blandin, 2020) to match series from the establishment and household surveys, respectively; weekly initial claims data from the BLS, whose true relationship with the monthly indicator is known to be one-to-one; the daily survey of consumers from Rasmussen, which is available since 2004 and for which we can estimate the relationship with the two monthly consumer surveys in our panel; and the Apple Mobility Trends data, which measures requests for driving directions on Apple devices, which we match to the historical series of U.S. vehicle miles traveled published by the Federal Highway Administration and add to our baseline panel.[24]

in the Gaussian case, so one alternative could be to tilt the mean forecast while preserving the uncertainty. Another possibility would be to run the regression with a very tight prior, and allow the posterior to move as months go by and sufficient history is accumulated for the novel indicator. With the exception of the Rasmussen series, for which a long enough history is available for the posterior to deviate meaningfully, we do not take either of these routes for simplicity of exposition.

[24]The weekly initial claims and bi-weekly Real-Time Population Survey are incorporated to our panel on the days in which they are originally published in real time. The Homebase employment data, the Chetty et al. (2020) credit card data and the Apple mobility trends were made public at various moments during the crisis, but we simulate the information set of an econometrician who would have had access to the data with one day delay since the beginning of the series.

5.2 Real-time tracking of the Great Lockdown

Figure 12: REAL-TIME ESTIMATE OF ECONOMIC ACTIVITY AROUND COVID-19 CRISIS

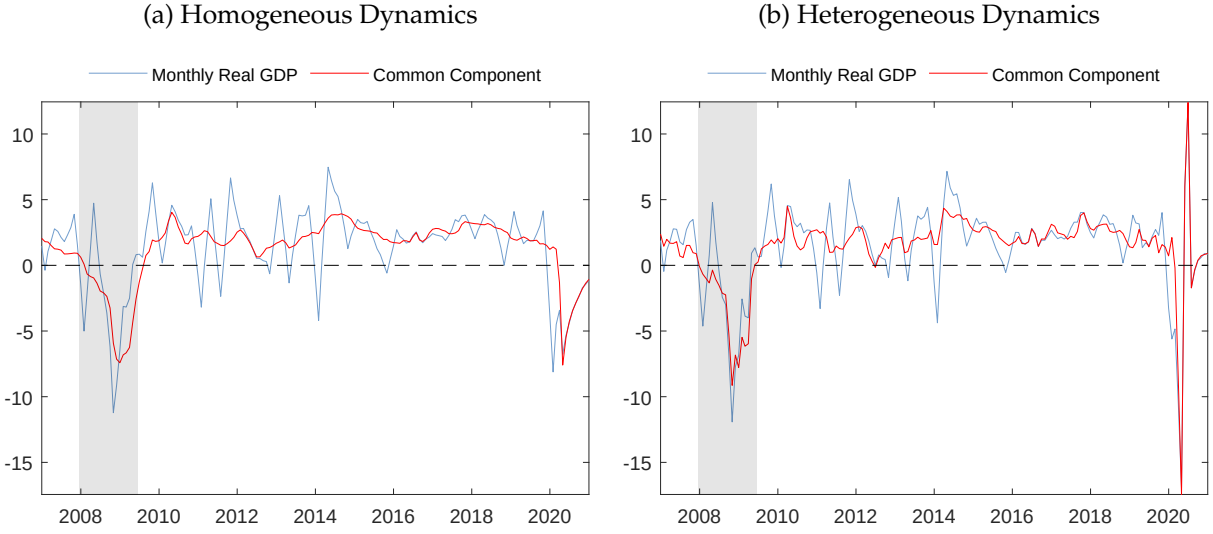


Notes. Daily estimate of US real GDP growth (blue line) with associated 68% HPD bands (green shaded areas). This version of the model conditions on novel data series as described in Section 5. The black line with circular markers is a version of the model with identical specification but which uses only the traditional monthly series. The figure illustrates how the incorporation of novel data leads to a more timely reading of the fall in activity. The vertical red lines indicate significant events. Gray shaded areas indicate NBER recessions.

Figure 12 displays the results, using our measure of rolling current-quarter GDP growth described above. It reproduces the right panel of Figure 2, but adds the mean estimate from a model estimated only with the traditional monthly data (black line with circular markers). As we can see, they both follow a similar path, but the model with alternative data is substantially earlier at capturing the decline in activity. Indeed, using only monthly data, the estimate of economic activity would only be slightly negative until the release of march industrial production and retail sales data on the 15th of April. Another interesting observation is how, as data for this period is released, the uncertainty around the estimates automatically increases as the model revises upwards the volatility of the common factor. Thus, not only activity is falling in an unprecedented

way, uncertainty is increasing massively. This contrasts with the results obtained with models without stochastic volatility, in which uncertainty always declines in response to new data, as discussed, e.g., by Banbura and Modugno (2014). After April 15, the two estimates are quite similar. We conclude that the alternative data were useful in capturing the decline in activity early but the model with traditional data tracked the rebound in a similar manner.

Figure 13: GDP FORECAST AND COMMON COMPONENT AS OF JUNE 1, 2020



Notes. Estimates of the path for real GDP growth and its common component using the vintage of data available on June 1, 2020. Panel (a) uses a version of the model with homogeneous dynamics ($m = 0$), whereas panel (b) is our full model with heterogeneous dynamics ($m = 1$).

Nevertheless, model specification still matters for the results. Figure 13 compares the estimates of the path for real GDP and its common component using the vintage of data available on June 1, 2020. Panel (a) uses a version of the model that imposes the traditional assumption of homogeneous dynamics ($m = 0$), whereas panel (b) is our full model with heterogeneous dynamics ($m = 1$). As can be seen, the results are qualitatively different. As discussed above, the homogeneous model places most weight on survey variables, and interprets the data as consistent with the US economy experiencing a large, persistent contraction similar to the 2008-2009 recession. The model, which places more weight on hard variables, instead foresees a much larger contraction in March and April, followed by a sharp rebound in May and June as the economy emerged from lockdown, consistent with an at least partial V-shaped recovery in the level of output. Ultimately, more time

is needed to formally evaluate the performance of different versions of the model when the final releases for GDP are published, but the message is clear: specification choices, can affect the results and the forecasts in qualitatively meaningful way even after incorporating novel sources of data.

6 Conclusion

We have proposed a Bayesian DFM, which incorporates low-frequency variation in the mean and variance of the variables, heterogeneous responses to common shocks, outlier observations and fat tails. In a comprehensive evaluation exercise based on fully real-time unrevised data, we have demonstrated that the real-time nowcasting performance is substantially improved across a variety of metrics. Capturing trends and SV improves nowcasting performance significantly and heterogeneous dynamics deliver substantial additional improvement. Fat tails successfully capture outlier observations in an automated way and are material to the model's behavior during the COVID-19 recession. Overall, this paper provides several advances to the nowcasting process.

References

ADRIAN, T., N. BOYARCHENKO, AND D. GIANNONE (2019): "Vulnerable growth," *American Economic Review*, 109, 1263–89.

ALVAREZ, R., M. CAMACHO, AND G. PEREZ-QUIROS (2012): "Finite sample performance of small versus large scale dynamic factor models," CEPR Discussion Papers 8867, C.E.P.R. Discussion Papers.

ANDREINI, P., C. IZZO, AND G. RICCO (2020): "Deep Dynamic Factor Models," Mimeo.

ANTOLIN-DIAZ, J., T. DRECHSEL, AND I. PETRELLA (2017): "Tracking the slowdown in long-run GDP growth," *Review of Economics and Statistics*, 99.

ANTOLIN-DIAZ, J., I. PETRELLA, AND J. F. RUBIO-RAMIREZ (2020): "Structural scenario analysis with SVARs," *Journal of Monetary Economics*.

ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): "Real-Time Measurement of Business Conditions," *Journal of Business & Economic Statistics*, 27, 417–427.

BAI, J. AND P. WANG (2015): "Identification and Bayesian Estimation of Dynamic Factor Models," *Journal of Business & Economic Statistics*, 33, 221–240.

BANBURA, M., D. GIANNONE, M. MODUGNO, AND L. REICHLIN (2012): "Now-Casting and the Real-Time Data Flow," Working Papers ECARES 2012-026, ULB.

BANBURA, M., D. GIANNONE, AND L. REICHLIN (2010): "Nowcasting," CEPR Discussion Papers 7883, C.E.P.R. Discussion Papers.

BANBURA, M. AND M. MODUGNO (2014): "Maximum Likelihood Estimation of Factor Models on Datasets with Arbitrary Pattern of Missing Data," *Journal of Applied Econometrics*, 29, 133–160.

BICK, A. AND A. BLANDIN (2020): "Real time labor market estimates during the 2020 coronavirus outbreak," *Unpublished Manuscript, Arizona State University*.

BLOOM, N. (2014): "Fluctuations in Uncertainty," *Journal of Economic Perspectives*, 28, 153–76.

BOIVIN, J. AND S. NG (2006): "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194.

BOK, B., D. CARATELLI, D. GIANNONE, A. M. SBORDONE, AND A. TAMBALOTTI (2018): "Macroeconomic nowcasting and forecasting with big data," *Annual Review of Economics*, 10, 615–643.

CARRIERO, A., T. E. CLARK, AND M. MASSIMILIANO (2020): "Nowcasting Tail Risks to Economic Activity with Many Indicators," Working Papers 202013R, Federal Reserve Bank of Cleveland.

CHETTY, R., J. N. FRIEDMAN, N. HENDREN, M. STEPNER, ET AL. (2020): "How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data," Tech. rep., National Bureau of Economic Research.

CIMADOMO, J., D. GIANNONE, M. LENZA, F. MONTI, AND A. SOKOL (2020): "Nowcasting with Large Bayesian Vector Autoregressions," .

CLARK, T. E. AND M. W. MCCRACKEN (2012): "In-sample tests of predictive ability: A new approach," *Journal of Econometrics*, 170, 1–14.

D'AGOSTINO, A., D. GIANNONE, M. LENZA, AND M. MODUGNO (2015): "Nowcasting Business Cycles: A Bayesian Approach to Dynamic Heterogeneous Factor Models," Tech. rep.

DIEBOLD, F. X. (2020): "Real-time real economic activity: Exiting the great recession and entering the pandemic recession," Tech. rep., National Bureau of Economic Research.

DIEBOLD, F. X. AND R. S. MARIANO (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–63.

DOAN, T., R. B. LITTERMAN, AND C. A. SIMS (1986): "Forecasting and conditional projection using realistic prior distribution," Staff Report 93, Federal Reserve Bank of Minneapolis.

DOZ, C., L. FERRARA, AND P.-A. PIONNIER (2020): "Business cycle dynamics after the Great Recession: An Extended Markov-Switching Dynamic Factor Model," PSE Working Papers halshs-02443364, HAL.

DOZ, C., D. GIANNONE, AND L. REICHLIN (2012): "A Quasi–Maximum Likelihood Approach for Large, Approximate Dynamic Factor Models," *The Review of Economics and Statistics*, 94, 1014–1024.

DURBIN, J. AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Methods: Second Edition*, Oxford University Press.

FAUST, J. AND J. H. WRIGHT (2013): "Forecasting Inflation," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, vol. 2.

FERNALD, J. (2014): "Productivity and Potential Output Before, During, and After the Great Recession," *NBER Macroeconomics Annual 2014*, 29.

FERNALD, J. G., J. H. STOCK, R. E. HALL, AND M. W. WATSON (2017): "The Disappointing Recovery of Output after 2009," *Brookings Papers on Economic Activity*.

FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2003): "Do financial variables help forecasting inflation and real activity in the euro area?" *Journal of Monetary Economics*, 50, 1243–1255.

GADEA-RIVAS, M. D., A. GOMEZ-LOSCOS, AND G. PEREZ-QUIROS (2014): "The Two Greatest. Great Recession vs. Great Moderation," WP Series 1423, Bank of Spain.

GEWEKE, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, North-Holland.

GIANNONE, D., L. REICHLIN, AND D. SMALL (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55, 665–676.

GROEN, J. J. J., R. PAAP, AND F. RAVAZZOLO (2013): "Real-Time Inflation Forecasting in a Changing World," *Journal of Business & Economic Statistics*, 31, 29–44.

HAMPEL, F. R., E. M. RONCHETTI, P. J. ROUSSEEUW, AND W. A. STAHEL (1986): *Robust statistics: the approach based on influence functions*, Probability and Mathematical Statistics Series, Wiley.

HARVEY, D., S. LEYBOURNE, AND P. NEWBOLD (1997): "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13, 281–291.

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185–212.

JURADO, K., S. C. LUDVIGSON, AND S. NG (2014): "Measuring Uncertainty," *American Economic Review, Forthcoming*.

KIM, C.-J. AND C. R. NELSON (1999): *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, The MIT Press.

KIM, S., N. SHEPHARD, AND S. CHIB (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–93.

KNOTEK, E. S. AND S. ZAMAN (2019): "Financial nowcasts and their usefulness in macroeconomic forecasting," *International Journal of Forecasting*, 35, 1708–1724.

LENZA, M. AND G. E. PRIMICERI (2020): "How to Estimate a VAR after March 2020," *Manuscript, Northwestern University*.

LEWIS, D. J., K. MERTENS, AND J. H. STOCK (2020): "Monitoring Real Activity in Real Time: The Weekly Economic Index," Liberty Street Economics 20200330b, Federal Reserve Bank of New York.

LUDVIGSON, S. C., S. MA, AND S. NG (2015): "Uncertainty and business cycles: exogenous impulse or endogenous response?" Tech. rep., National Bureau of Economic Research.

MARCELLINO, M., M. PORQUEDDU, AND F. VENDITTI (2014): "Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility," *Journal of Business & Economic Statistics, Forthcoming*.

MARIANO, R. S. AND Y. MURASAWA (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

MCCONNELL, M. M. AND G. PEREZ-QUIROS (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?" *American Economic Review*, 90, 1464–1476.

MOENCH, E., S. NG, AND S. POTTER (2013): "Dynamic hierarchical factor models," *Review of Economics and Statistics*, 95, 1811–1817.

PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–852.

PRIMICERI, G. E. AND A. TAMBALOTTI (2020): "Macroeconomic Forecasting in the Time of COVID-19," *Manuscript, Northwestern University*.

SARGENT, T. J. AND C. A. SIMS (1977): "Business cycle modeling without pretending to have too much a priori economic theory," Working Papers 55, Federal Reserve Bank of Minneapolis.

SCHORFHEIDE, F., D. SONG, ET AL. (2020): "Real-time forecasting with a (standard) mixed-frequency var during a pandemic," Tech. rep.

SIMS, C. A. (2012): "Comment to Stock and Watson (2012)," *Brookings Papers on Economic Activity*, Spring, 81–156.

STOCK, J. AND M. WATSON (2012): "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity*, Spring, 81–156.

STOCK, J. H. AND M. W. WATSON (1989): "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual 1989, Volume 4*, National Bureau of Economic Research, Inc, NBER Chapters, 351–409.

——— (2003): "Has the Business Cycle Changed and Why?" *NBER Macroeconomics Annual 2002*, 17, 159–230.

——— (2005): "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Papers 11467, National Bureau of Economic Research, Inc.

Online Appendix to "Nowcasting Economic Activity with Secular Trends, Large Shocks and Alternative Data"

by Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

Contents

A Treatment of the missing observations in the vectorized Kalman Smoother

To be completed.

B Details on model and algorithm

B.1 Details of the Gibbs sampler algorithm

To be completed.

B.2 Procedure to set the scale of the priors

To be completed.

B.3 Data Series included in the analysis

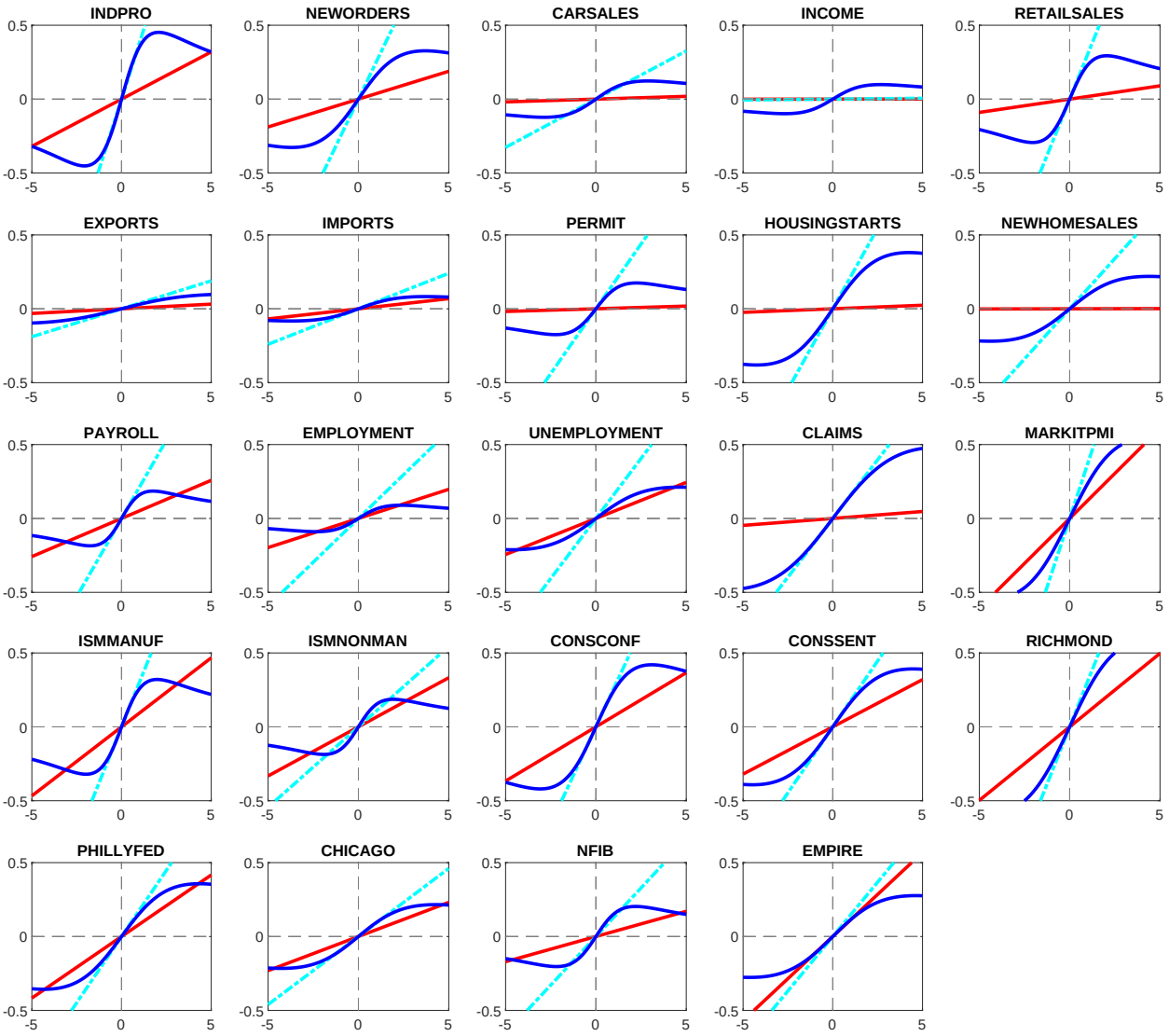
Table B.1: DATA SERIES USED FOR US EMPIRICAL ANALYSIS

	Type	Start Date	Transform.	Lag
QUARTERLY TIME SERIES				
Real GDP	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real GDI	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Consumption (excl. durables)	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Investment (incl. durable cons.)	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Total Hours Worked	Labor Market	Q2:1948	% QoQ Ann	28
MONTHLY INDICATORS				
Real Personal Income less Transfers	Expenditure & Inc.	Feb 59	% MoM	27
Industrial Production	Production & Sales	Jan 47	% MoM	15
New Orders of Capital Goods	Production & Sales	Mar 68	% MoM	25
Real Retail Sales & Food Services	Production & Sales	Feb 47	% MoM	15
Light Weight Vehicle Sales	Production & Sales	Feb 67	% MoM	1
Real Exports of Goods	Foreign Trade	Feb 68	% MoM	35
Real Imports of Goods	Foreign Trade	Feb 69	% MoM	35
Building Permits	Housing	Feb 60	% MoM	19
Housing Starts	Housing	Feb 59	% MoM	26
New Home Sales	Housing	Feb 63	% MoM	26
Payroll Empl. (Establishment Survey)	Labor Market	Jan 47	% MoM	5
Civilian Empl. (Household Survey)	Labor Market	Feb 48	% MoM	5
Unemployed	Labor Market	Feb 48	% MoM	5
Initial Claims for Unempl. Insurance	Labor Market	Feb 48	% MoM	4
MONTHLY INDICATORS (SOFT)				
Markit Manufacturing PMI	Business Confidence	May 07	-	-7
ISM Manufacturing PMI	Business Confidence	Jan 48	-	1
ISM Non-manufacturing PMI	Business Confidence	Jul 97	-	3
NFIB Small Business Optimism Index	Business Confidence	Oct 75	Diff 12 M.	15
U. of Michigan: Consumer Sentiment	Consumer Confid.	May 60	Diff 12 M.	-15
Conf. Board: Consumer Confidence	Consumer Confid.	Feb 68	Diff 12 M.	-5
Empire State Manufacturing Survey	Business (Regional)	Jul 01	-	-15
Richmond Fed Mfg Survey	Business (Regional)	Nov 93	-	-5
Chicago PMI	Business (Regional)	Feb 67	-	0
Philadelphia Fed Business Outlook	Business (Regional)	May 68	-	0

Notes. % QoQ Ann refers to the quarter on quarter annualized growth rate, % MoM refers to $(y_t - y_{t-1})/y_{t-1}$ while Diff 12 M. refers to $y_t - y_{t-12}$. The last column shows the average publication lag, i.e. the number of days elapsed from the end of the period that the data point refers to until its publication by the statistical agency. All series were obtained from the Haver Analytics database.

C Additional results

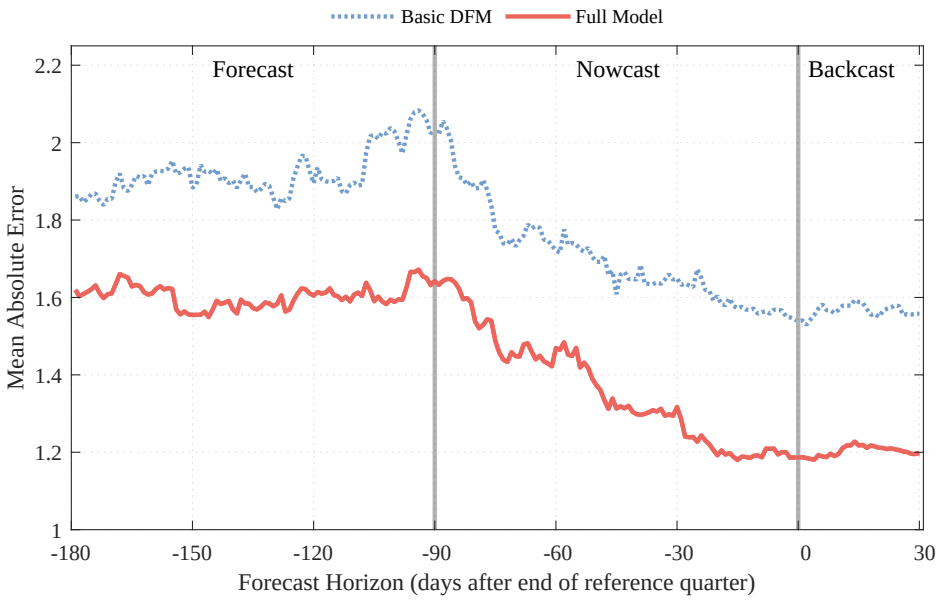
Figure C.1: INFLUENCE FUNCTIONS FOR ALL VARIABLES



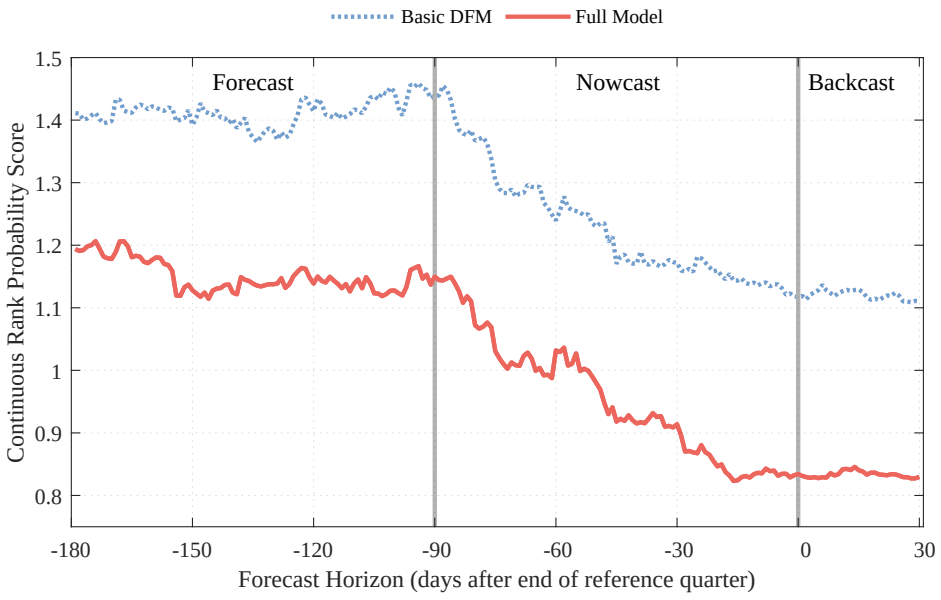
Notes. The panels of the figure plot the *influence functions* for different variables, that is, by how much the estimate of the dynamic factor is updated when the release in the variable is different from its forecast and thus contains "news". The blue lines plot these influence functions in the Gaussian case while the red lines represent the Student-t case. As shown in equations (13) - (16) the model with fat tails allows these functions to be nonlinear and nonmonotonic.

Figure C.2: FORECAST EVALUATION AS THE DATA FLOW ARRIVES (ALTERNATIVE METRICS)

(a) Point forecast accuracy: MAE (accurate forecast implies low value)



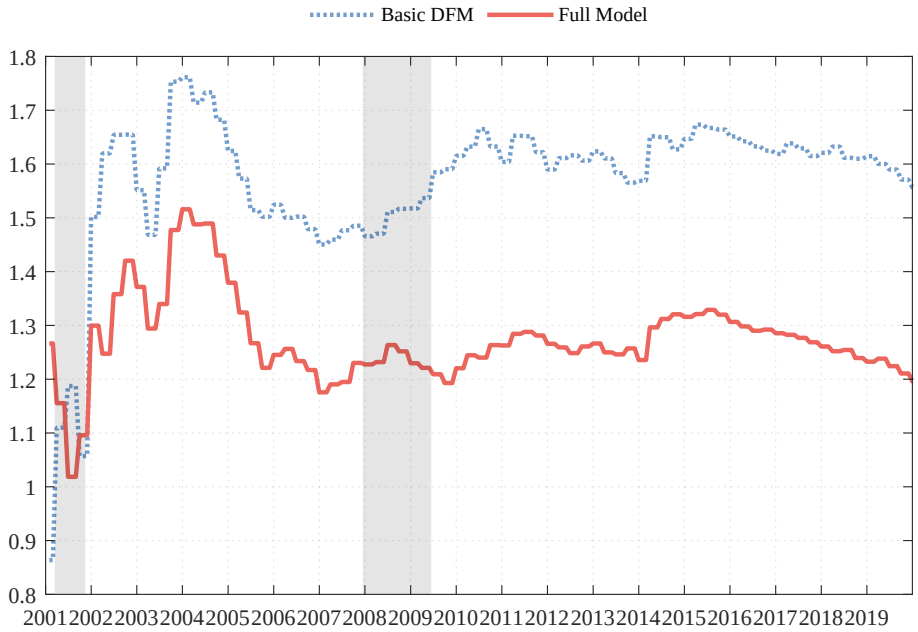
(b) Density forecast accuracy: CRPS (accurate forecast implies low value)



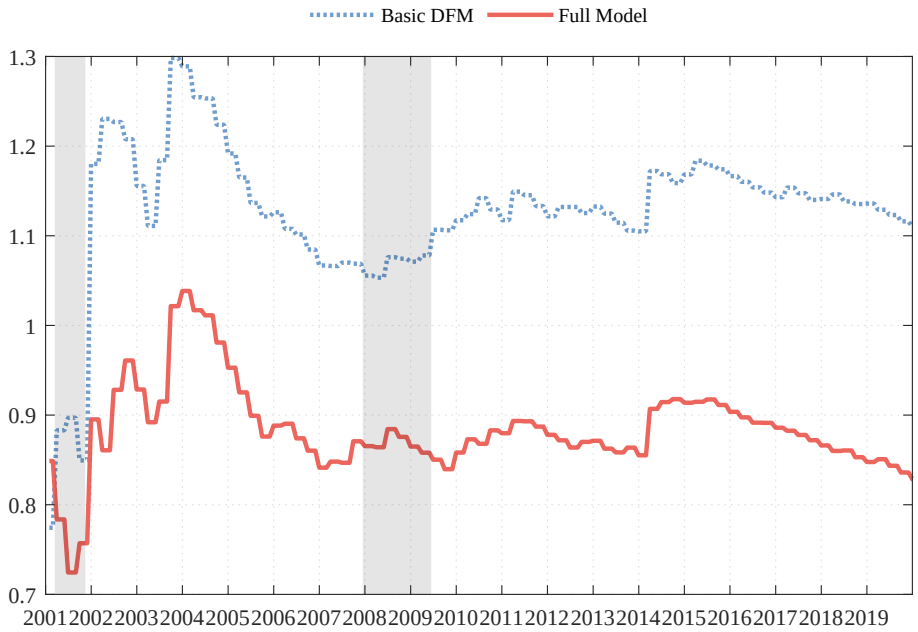
Notes. In both panels, the horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter of a given GDP release. Thus, from the point of view of the GDP forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of the next quarter; forecasts at a 90-0 day horizon are nowcasts of the current quarter, and the forecasts produced 0-25 days after the end of the quarter are backcasts of the last quarter. Panel (a) plots the MAE of the full model (solid red line) as well as the basic DFM (dotted blue) over this horizon. A more accurate forecasts implies a lower MAE. Panel (b) displays the analogous evolution of the CRPS of the two models, a measure of density forecast accuracy, which takes a lower value for a more accurate forecast. The differences between the models are statistically significant throughout the horizon in both panels.

Figure C.3: FORECAST EVALUATION THROUGH TIME (ALTERNATIVE METRICS)

(a) Point forecast accuracy: MAE (accurate forecast implies low value)



(b) Density forecast accuracy: CRPS (accurate forecast implies low value)



Notes. In both panels, the horizontal axis indicates the time of the evaluation sample 2000-2019. Panel (a) plots the rolling MAE of the full model (solid red line) as well as the basic DFM (dotted blue) through. A more accurate forecast implies a lower MAE. Panel (b) displays the analogous rolling evolution of the CRPS of the two models, a measure of density forecast accuracy, which also takes a lower value for a more accurate forecast. In both panels, the gray shaded areas indicate NBER recessions. The rolling metrics indicate that the full DFM is preferred based on both point and density forecasting performance as early as 2002.