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*

June 14, 2021

Abstract

A key question for households, firms, and policy makers is: how is the economy doing now? We develop a Bayesian dynamic factor model and compute daily estimates of US GDP growth. Our framework gives prominence to features of modern business cycles absent in linear Gaussian models, including movements in long-run growth, time-varying uncertainty, and fat tails. We also incorporate newly available high-frequency data on consumer behavior. The model beats benchmark econometric models and survey expectations at predicting GDP growth over two decades, and advances our understanding of macroeconomic data during the COVID-19 pandemic.

Keywords: Nowcasting; Daily economic index; Dynamic factor models; Real-time data; Bayesian Methods; Fat Tails; Covid-19 Recession.

JEL Classification Numbers: E32, E23, O47, C32, E01.

*We thank seminar participants at the NBER-NSF Seminar in Bayesian Inference in Econometrics and Statistics, the ASSA Big Data and Forecasting Session, the World Congress of the Econometric Society, the Barcelona Summer Forum, the Bank of England, Bank of Italy, Bank of Spain, Bank for International Settlements, Norges Bank, the Banque de France PSE Workshop on Macroeconometrics and Time Series, the CFE-ERCIM in London, the DC Forecasting Seminar at George Washington University, the EC2 Conference on High-Dimensional Modeling in Time Series, Fulcrum Asset Management, IHS Vienna, the IWH Macroeconometric Workshop on Forecasting and Uncertainty, and the NIESR Workshop on the Impact of the Covid-19 Pandemic on Macroeconomic Forecasting. We are very grateful to Boragan Aruoba and Domenico Giannone for extensive discussions. We also especially thank 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, and Chris Adjaho, Alberto D'Onofrio, and Yad Selvakumar for outstanding research assistance. 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

Assessing macroeconomic conditions in real time is challenging. The most important indicators, such as Gross Domestic Product (GDP), are published only on a quarterly basis and with considerable delay, while many 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. On top of these peculiarities of the data collection and publication process, the first two decades of this century have made clear that secular change, evolving uncertainty and large shocks are pervasive features of modern macroeconomic fluctuations.

This paper advances macroeconomic "nowcasting" by proposing a novel Bayesian dynamic factor model (DFM) that explicitly incorporates these features. We show that our model outperforms benchmark statistical models at real-time predictions of GDP growth, and improves upon survey expectations of professional forecasters. Modeling long-run growth as time-varying eliminates a bias often present in long-horizon forecasts. Moreover, our explicit treatment of non-linearities and fat tails is important throughout the postwar period; it becomes critical to track economic activity in crisis periods, such as the Great Recession of 2008-2009, and most notably the COVID-19 pandemic during which existing macroeconometric models have struggled to produce useful results (Lenza and Primiceri, 2020). We also propose a method to incorporate newly available high-frequency data on consumer behaviour, such as credit card transactions (Chetty et al., 2020), driving trips, or restaurant reservations, into DFMs. Our analysis highlights that the new data are helpful for tracking activity during the pandemic, but that a careful econometric specification is just as important.

DFMs are based on the idea that the variation in a large number of macroeconomic time series is driven by a small number of shocks, whose propagation can be summarized by a few common factors (Sargent and Sims, 1977; Stock and Watson, 1989;

Forni et al., 2003). Our Bayesian DFM is estimated on 29 key US macroeconomic time series, and incorporates low-frequency variation in the mean and variance of the series, heterogeneous lead-lag patterns in the responses of variables to common shocks, and fat tails. Figure 1 illustrates the importance of these features in US data. Panel (a) plots the annual growth rate of US GDP, together with its mean and standard deviation in selected subsamples. It highlights the presence of secular economic trends such as the recent stagnation in long-term growth or the "Great Moderation" in macroeconomic volatility. Panel (b) compares the annual growth rate of selected indicators to that of GDP. It makes clear that lead-lag dynamics are a systematic feature of the data, with investment and durable spending leading GDP and labor variables lagging it. Panel (c) illustrates how large, one-off outliers in individual series, typically in the level of the series, and usually caused by tax changes, strikes or natural disasters, are endemic to macroeconomic data. Despite their prominent role, all of the above features of the data are mostly assumed away in existing macroeconometric models. In this paper we show that introducing them explicitly into a DFM improves our understanding of economic developments in real time.[1] Our methodological contribution is the development of a fast and efficient algorithm for posterior simulation, which can handle non-linearities and non-Gaussianities without losing the intuition and computational ease of Kalman Filtering and Gibbs sampling. Our Bayesian methods give an important role to probabilistic assessments of economic conditions, a departure from existing approaches which favor classical estimation techniques and focus on point forecasts.

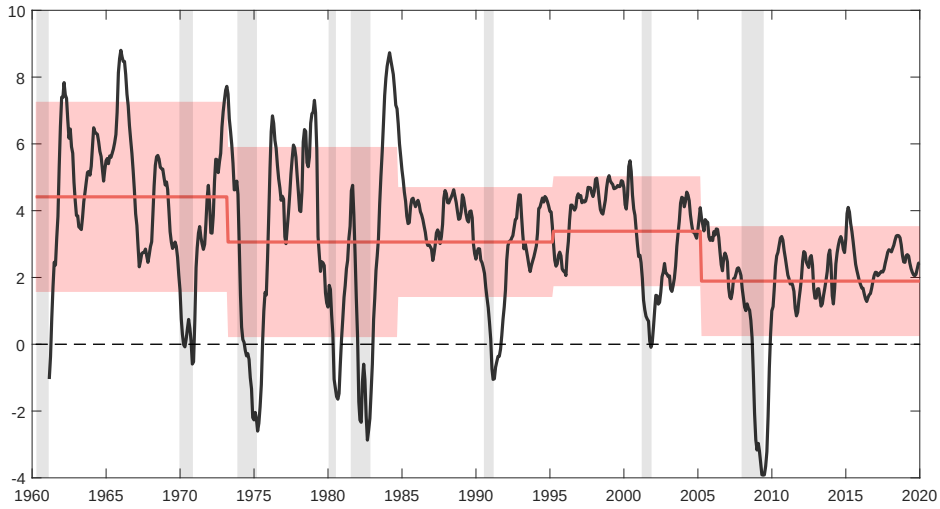
Beyond our methodological contribution, we provide three empirical contributions. First, we construct daily estimates of US real GDP growth covering January 2000 to August 2020. To this end, we build a new real-time database which is updated every day in which data are released, and re-estimate the model with each update.[2] Figure

[1] While Figure 1 focuses on the US, instabilities in the data of the sort we are modeling in our framework are even more likely to be a challenge for nowcasting in other countries, especially emerging and developing economies.

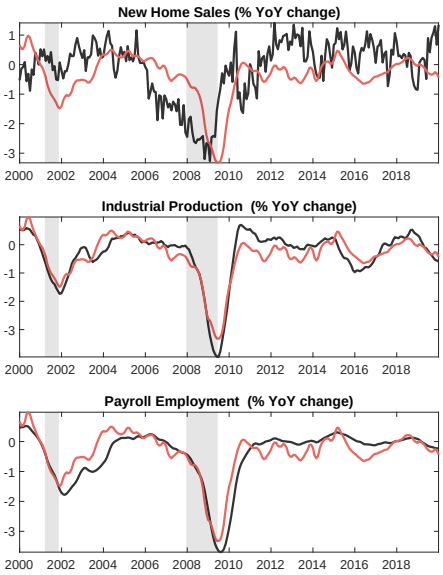
[2] Despite the efficiency of our Monte Carlo Markov Chain algorithm, the sheer scale of such an exercise would be infeasible using standard computer power. We enable it through the use of modern cloud computing.

Figure 1: SALIENT FEATURES OF THE MACROECONOMIC DATA FLOW

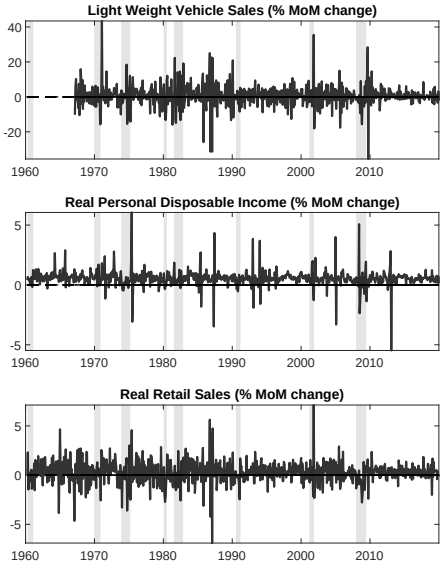
(a) Time-varying mean and volatility of US real GDP growth



(b) Heterogeneous lead/lag dynamics



(c) Fat tails in macroeconomic data



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 over selected subsamples, with meaningful changes visible across the subsamples. Panel (b) plots the twelve-month growth rate of selected indicators of economic activity (black) together with four-quarter real GDP growth (red). This illustrates how the business-cycle comovement of macroeconomic variables features heterogeneous patterns of leading and lagging dynamics. Panel (c) presents raw data series for selected indicators of economic activity. These highlight the presence of fat-tailed outliers in macroeconomic time series. The gray shaded areas indicate NBER recessions.

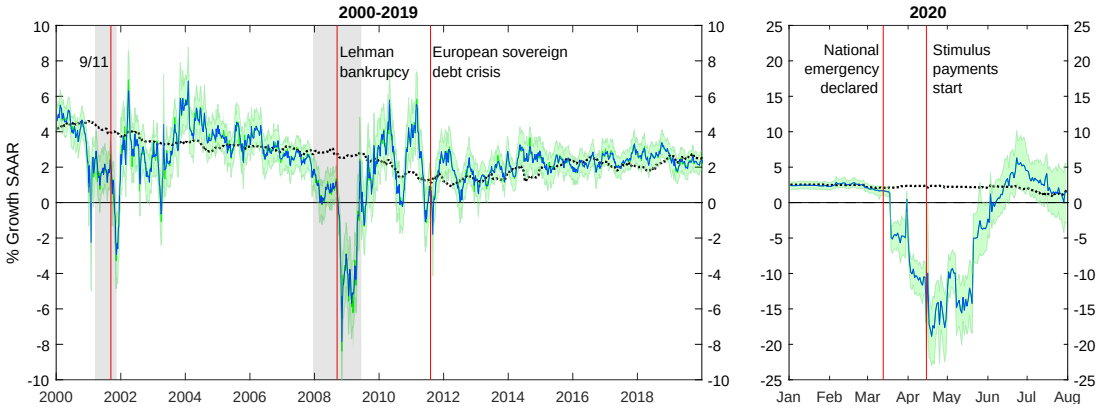
2 presents the resulting daily estimate of current-quarter GDP growth, together with the estimated uncertainty around it, using information only up to each point in time. As the data arrive, the model tracks in real time important developments of the last two decades. We formally evaluate the accuracy of these estimates over the period 2000-2019, by comparing the nowcasts of the model with the official data subsequently published by the Bureau of Economic Analysis. We find that our model delivers a large and highly significant improvement relative to various model benchmarks: a basic DFM, and the model maintained by the New York Fed. Both point and density forecasting improve in a statistically and economically significant manner, and each of the new model components adds to the improvement. We also find that our model's nowcasts are more accurate than 80% of individual panelists from the Survey of Professional Forecasters (SPF), and indistinguishable from the survey median. A comparison with the Federal Reserve staff Greenbook projections reveals that early in quarter, our model forecasts are as accurate as the Greenbook's, with the latter taking an advantage later on.[3] We thus document, contrary to earlier results by Faust and Wright (2009), that short-horizon forecasts from state-of-the-art econometric models, private-sector survey expectations, and the Fed's Greenbook, can be competitive with each other, though the Fed appears to display an informational advantage (Romer and Romer, 2000; Nakamura and Steinsson, 2018). At longer horizons, both SPF surveys and FOMC projections display a noticeable upward bias in the last decade, which our model eliminates thanks to the addition of time-varying long-run growth.

Our second empirical contribution is to use the model to study in detail the developments in the US economic during 2020, the period surrounding the COVID-19 pandemic. The dramatic speed and size of the recent recession poses unique challenges to macroeconometric models. The new components of the model, in combination, are

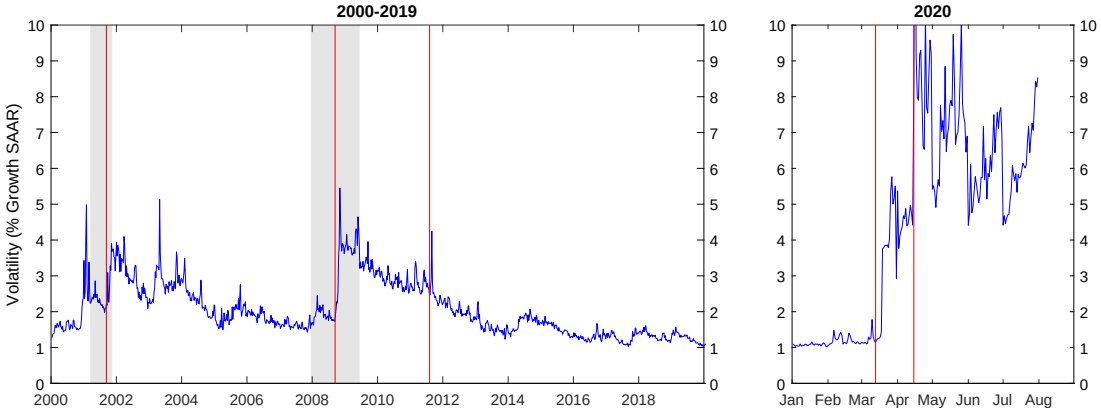
[3]Human forecasters are a high benchmark for econometric models in particular for short-run forecasts, as they have access to a large information set including news, political developments, and information about structural change. Consensus measures, such as the SPF median and institutional forecasts, are an even higher benchmark as they aggregate the opinions of a multitude of forecasters (Sims, 2002).

Figure 2: DAILY REAL-TIME ESTIMATES: 2000-2019 VS. 2020

(a) Current-quarter US real GDP growth, daily assessment



(b) Volatility of current-quarter growth, daily assessment



Notes. Panel (a) plots the daily estimate of US real GDP growth (blue line) with associated uncertainty (green shaded areas) computed from our Bayesian DFM. The left chart plots this estimate from 2000 through to the end of 2019, the right chart for the period January-August 2020 (note the different scale). The vertical red lines indicate significant events and gray shaded areas indicate NBER recessions. Panel (b) provides an analogous chart for the volatility of the daily activity estimate in Panel (a).

critical to obtain a timely reading of the US economy during the Great Lockdown. Two important insights are that fat-tailed observations were a pervasive feature of macroeconomic time series even before the pandemic, and that the COVID-19 recession is more than just a "macroeconomic outlier." Instead, through the lens of the model the COVID-19 recession is a massive drop in aggregate activity, a large increase in overall uncertainty, and a strong departure from the usual business cycle comovement

patterns of investment, employment, and consumption. Specifically, the lockdown produced outsized contractions in sectors such as services which normally display little cyclicality. Moreover, the aggressive policy response eliminated the typical comovement between different variables, such as filings for unemployment insurance and aggregate consumption which are usually strongly negatively correlated. This is problematic for factor models, which rely on the average comovement between different series to estimate the factors. Our model can distinguish between persistent increases in aggregate volatility, and transitory ones that are specific to individual indicators. It is therefore able to separate the aggregate shock from large, purely transitory outliers which are detected only on a handful of series directly impacted by the lockdown. It can also capture different shapes of recoveries, in line with the divergent behavior of durable and non-durable spending recently highlighted by Beraja and Wolf (2021). For the COVID-19 recession, it tracks an unprecedented fall in activity in 2020, followed by an unusual rebound which is quick but partial.

Finally, in response to the pandemic a number of high-frequency data on consumer behavior, such as credit card transactions (Chetty et al., 2020), driving trips, or restaurant reservations, have become available. These "alternative data" offer the promise of a more timely reading of economic conditions than the traditional monthly indicators, and observers in media and policy circles pay close attention to them. Each of them however provides only a partial and noisy signal of underlying conditions, so it is important to exploit their *joint* information in a systematic way, and quantify their contribution to our assessment of the economy. This was precisely the motivation behind the development of factor models (Geweke, 1977; Stock and Watson, 1989). A key challenge is that many of the new series have extremely short histories, sometimes as short as two months, complicating econometric inference. To overcome this, we take a Bayesian approach to condition the yet-to-be-released 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 use information on credit card transactions or cell phone mobility data to produce updated estimates of aggregate consumption or vehicle miles traveled, both of which are part of the panel on which we estimate our DFM. We find that the information contained in the newly available series contributes to a more timely reading of economic conditions in 2020, but the model's ability to handle changes in volatility and fat tails is just as important. Since the COVID-19 episode will remain part of macroeconomic times series, in the same way that a binding Zero Lower Bound on interest rates became a permanent feature of the data after 2008, the introduction of nonlinearities and fat tails into empirical macro models will become pressing even as the pandemic moves into the rear-view mirror.

Contribution to the literature. Methodologically, our work advances the literature that models macroeconomic time series with DFMs. Important contributions include Stock and Watson (2002a,b) and Aruoba, Diebold, and Scotti (2009). Giannone, Reichlin, and Small (2008) formalized the application of DFMs to nowcasting.[4] Our paper gives a prominent role to what Sims (2012) calls "recurrent phenomena" of macroeconomic time series that the DFM 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. In Antolin-Diaz, Drechsel, and Petrella (2017) we made the case for allowing both shifts in long-run growth and changes in volatility. Here we develop the DFM framework further to include heterogeneous dynamics and student-t distributed outliers, and study in detail their importance for the nowcasting problem. The present paper is the first to investigate the introduction of student-t distributed outliers within DFMs.[5] Moreover, in contrast to the majority of the literature, we take a Bayesian perspective to estimation, and give emphasis to

[4] See also Aruoba and Diebold (2010), and Banbura et al. (2013) and Stock and Watson (2017) for useful surveys.

[5] Doz, Ferrara, and Pionnier (2020) also emphasize the importance of low-frequency trends. Marcellino, Porqueddu, and Venditti (2016) were the first to introduce time varying volatility in DFMs, whereas Camacho and Perez-Quiros (2010) and D'Agostino, Giannone, Lenza, and Modugno (2016) stress the importance of accounting for the asynchronous nature of macroeconomic data within a DFM.

probabilistic prediction and real-time density forecasts. More broadly, we contribute to an empirical literature that stresses the importance of modelling time-variation, non-linearities and departures from normality in macroeconomic models, including Vector Autoregressions (VARs) and Dynamic Stochastic General Equilibrium (DSGE) models. See, e.g. Cogley and Sargent (2005), Primiceri (2005), Cúrdia, Del Negro, and Greenwald (2014), Fernández-Villaverde et al. (2015), or Brunnermeier et al. (2020).

Finally, this paper tackles the challenge of time series modeling with macroeconomic data from 2020 and thereafter. Other attempts in this direction include Primiceri and Tambalotti (2020), Lenza and Primiceri (2020), Schorfheide and Song (2020), and Cimadomo et al. (2020), all of which discuss the challenges to using VAR models in 2020. Diebold (2020) studies the performance of the Aruoba, Diebold, and Scotti (2009) model during the pandemic. We propose a framework that helps us interpret the 2020 data, and show that prominent features of the data during this period, such as time-varying volatility and outliers, were already critical to understanding economic activity in the previous two decades. Lewis et al. (2020) highlight the usefulness of using weekly data for macroeconomic monitoring during the pandemic. Chetty et al. (2020) propose to use newly available high-frequency data to track the economy in real time. To the best of our knowledge, we are the first to systematically integrate these time series into the DFM framework.

Structure of the paper. Section 2 introduces the econometric framework. 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 for 2000-2019 are presented in Section 4. Section 5 investigates the COVID-19 recession and recovery through the lens of our model and analyzes the importance of incorporating newly available high-frequency data for the nowcasting process. Section 6 concludes.

2 Econometric framework

2.1 The dynamic factor 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 display outliers, denoted $\mathbf{o}_t$. Formally,

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

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, is applied to $\mathbf{y}_t - \mathbf{o}_t$, which makes clear that the *level* of the variables displays outliers, while the factor structure is present in the growth rates.[6] This captures the fact that 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, with the series returning to its original level once their effect dissipates. This often leads to consecutive outliers with opposite sign in the growth rate (see Panel (c) of Figure 1).

Low-frequency changes in the long-run growth rate of $\mathbf{y}_t$ are captured by time-variation in $\mathbf{c}_t$. One could allow time-varying intercepts in all or a subset of the variables in the system, and 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. In our application, we specify $\mathbf{c}_t$ as

$$\mathbf{c}_t = \mathbf{c} + \mathbf{b}a_t, \tag{2}$$

[6] We measure $\Delta\mathbf{y}_t$ in the data as the percentage change, i.e. $100 \times (y_t - y_{t-1})/y_{t-1}$. Some variables, e.g. surveys, are stationary in levels, so the difference operator would not apply to them.

where a_t is a common time-varying long-run growth rate, $\mathbf{b}$ is a selection vector taking value 1 for the series representing the expenditure and income measures of GDP, as well as consumption, and 0 for all other variables. $\mathbf{c}$ is a vector of constants.[7] We estimate the model on a panel of real activity variables and focus on the case of a single factor ($k = 1$, $\mathbf{f}_t = f_t$). The relevant laws of motion are specified as

$$
\begin{aligned}
(1 - \phi(L))f_t &= \sigma_{\varepsilon_t}\varepsilon_t, && (3)\\
(1 - \rho_i(L))u_{i,t} &= \sigma_{\eta_{i,t}}\eta_{i,t}, \qquad i = 1, \ldots, n && (4)\\
o_{i,t} &\overset{iid}{\sim} t_{v_i}(0, \sigma^2_{o,i}), \quad i = 1, \ldots, n && (5)
\end{aligned}
$$

where $\phi(L)$ and $\rho_i(L)$ denote polynomials in the lag operator of orders p and q. The idiosyncratic components are cross-sectionally orthogonal and uncorrelated with the common factor at all horizons, i.e. $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and $\eta_{i,t} \overset{iid}{\sim} N(0,1)$. The outliers are modeled as independent additive Student-t innovations, with scale and the degrees of freedom, $\sigma_{o,i}$ and v_i, to be estimated jointly with the other parameters of the model. The fat-tailed components are independent from the factor and idiosyncratic innovations. Following Primiceri (2005), the time-varying parameters are driftless random walks:

$$
\begin{aligned}
a_t &= a_{t-1} + v_{a,t}, && v_{a,t} \overset{iid}{\sim} N(0, \omega_a^2) && (6)\\
\log \sigma_{\varepsilon_t} &= \log \sigma_{\varepsilon_{t-1}} + v_{\varepsilon,t}, && v_{\varepsilon,t} \overset{iid}{\sim} N(0, \omega_\varepsilon^2) && (7)\\
\log \sigma_{\eta_{i,t}} &= \log \sigma_{\eta_{i,t-1}} + v_{\eta_{i,t}}, && v_{\eta_{i,t}} \overset{iid}{\sim} N(0, \omega_{\eta,i}^2) \quad i = 1, \ldots, n && (8)
\end{aligned}
$$

where σ_{ε_t} and $\sigma_{\eta_{i,t}}$ capture the stochastic volatility (SV) of the innovations to factor and idiosyncratic components.

[7] Antolin-Diaz et al. (2017) study the specification of trends in DFMs, and find that the long-run growth rates of interest are retrieved correctly even if any low frequency component of a remaining variable is incorrectly specified as a constant. This is also true if the long-run components are shared with GDP.

2.2 Dealing with mixed frequencies and missing data

The model is specified at monthly frequency. Following Mariano and Murasawa (2003), the (observed) growth rates of quarterly variables, x_t^q, are linked to the (unobserved) monthly growth rate x_t^m, where every third observation of x_t^q is observed, by

$$x_t^q = \frac{1}{3}x_t^m + \frac{2}{3}x_{t-1}^m + x_{t-2}^m + \frac{2}{3}x_{t-3}^m + \frac{1}{3}x_{t-4}^m. \tag{9}$$

This reduces the presence of mixed frequencies to a problem of missing data in a monthly model.[8] 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.

2.3 Priors and model settings

The number of lags in $\mathbf{\Lambda}(\mathbf{L})$, $\phi(L)$, and $\rho_i(L)$ 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. 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 our 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 tradition 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 $\phi(L)$ 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 for the contemporaneous coefficient is set to $\hat{s}_i$, an estimate of the standard deviation of each of the variables,

[8] Additional sources of missing data include the "ragged edge" at the sample end coming from the non-synchronicity of data releases, and missing data at the beginning of the sample for more recent series.

and to zero in subsequent lags. This prior reflects the belief that the factor is a cross-sectional average of the standardized variables, see D'Agostino et al. (2016). For the autoregressive coefficients of the idiosyncratic components, $\rho_i(L)$ the prior is set to zero for all lags, shrinking 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. 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 a DFM without time-varying long-run 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.[9] For v_i, we use a weakly informative prior specified as a Gamma distribution $\Gamma(2, 10)$ discretized on the support [3; 40]. The lower bound at 3 enforces the existence of a conditional variance.

2.4 Estimation algorithm

We build a highly efficient Gibbs sampler algorithm to draw from the joint posterior distribution of the parameters and latent objects, adapting ideas from Moench et al. (2013). The SV step follows 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 Student-t 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 a hierarchical algorithm that avoids large state-spaces and allows parallelization of many steps, leading to extremely fast computation. In addition, we implement a vectorized version

[9] Antolin-Diaz et al. (2017) provide a number of robustness checks around the choice of priors in a Bayesian DFM.

of the Kalman filter/smoother (sometimes referred to as precision sampler, see Chan and Jeliazkov, 2009), which we extend to allow for missing observations (see Online Appendix A). The vectorized approach leads to substantial gains in computational time. The Online Appendix presents the details of our algorithm. Below we provide a sketch.

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* $\mathbf{c}_t, \mathbf{o}_t$, *and* f_t, *are initialized by running the Kalman filter and smoother conditional on the initial values of the parameters.*
2. *For each variable,* $i = 1, \ldots, N$*:*
   - 2.1. *Compute* $\Delta y_{i,t} - c_{i,t} - \lambda_i(L) f_t = u_{i,t} + \Delta o_{i,t}$.
   - 2.2. *Use the Kalman filter and simulation smoother to independently draw the Student-t components. Draw the associated parameters* $\sigma_{o,i}$, v_i.
   - 2.3. *Compute outlier adjusted variable* $\Delta y_{i,t}^{OA} = \Delta(y_{i,t} - o_{i,t}) = c_{i,t} + \lambda_i(L) f_t + u_{i,t}$.
3. *Conditional on the outlier adjusted variables:*
   - 3.1. *Use the Kalman filter and simulation smoother to draw* $\{\mathbf{c}_t, f_t\}$.
   - 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, and 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.*

We note several points. First, the algorithm iterates between a univariate state space 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 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, so it can be run in parallel using multi-core processors. Finally, 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 (see Bai and Wang, 2015).

2.5 Variable selection

Our DFM includes variables measuring US real economic activity, excluding prices, monetary and financial variables, and is specified with a single common factor.[10] We include all of the available surveys of consumer and business sentiment. The timeliness of these data series (they are released around the end of each month, referring to conditions prevailing during the month) make them especially valuable for nowcasting. We do not use disaggregated data (e.g. sector-level production measures) and rely only on the headline indicators for each category.[11] In total, we include 29 time series, which are listed with detailed explanations in Online Appendix B.

3 General implications for tracking economic activity

We discuss the implications of our modeling assumptions for tracking US economic activity as new data arrive. We highlight the importance of capturing slow-moving trends in long-run growth and volatility, the evidence for heterogeneity in the responses of macro variables to common shocks, and the implications of fat-tailed observations for updating beliefs about the state of the economy. This is an in-sample analysis that excludes the COVID-19 episode, which we cover in Section 5.

3.1 Secular trends: long-run growth and drifting volatilities

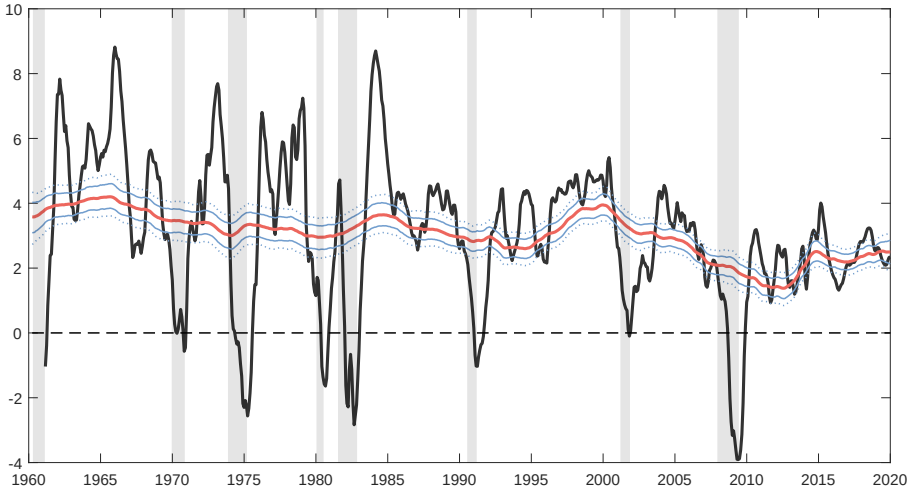
Figure 3 displays the estimate of the time-varying long-run growth rate of GDP as well as the SV of the common factor. These estimates condition on the full sample

[10] Giannone et al. (2008) conclude that that prices and monetary indicators do not improve GDP nowcasts. Banbura et al. (2013), Forni et al. (2003) and Stock and Watson (2003) find mixed results for financial indicators.

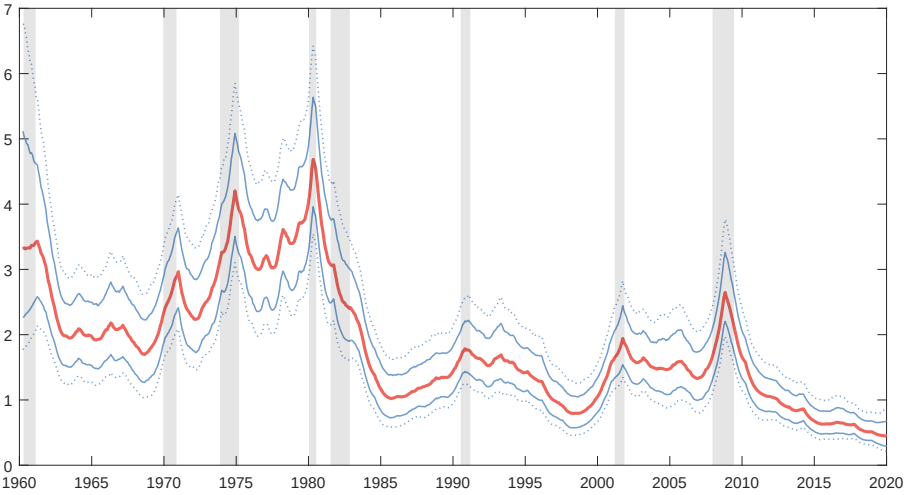
[11] Banbura et al. (2013), among others, argue that the strong correlation in the idiosyncratic components of disaggregated series of a given category results in misspecification that can worsen the in-sample fit and out-of-sample forecasting performance of DFMs.

Figure 3: ESTIMATED LOW-FREQUENCY COMPONENTS OF US GDP GROWTH

(a) Real GDP growth and long-run trend



(b) Volatility of the common activity factor



Notes. Panel (a) plots the growth rate of US real GDP (solid black line), the posterior median (solid red) and the 68% and 90% (solid and dotted blue) High Posterior Density intervals of the time-varying long-run growth rate. 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) High Posterior Density credible intervals of the volatility of the common factor. Our estimate implies a trend decline in volatility (Great Moderation), as well as heightened volatility in economic contractions. Gray shaded areas represent NBER recessions.

and account for both filtering and parameter uncertainty. Panel (a) presents the posterior median and the 68% and 90% high posterior density (HPD) intervals of the long-run growth rate of US real GDP, together with the raw data series for real GDP growth. The estimate from our model conforms with the established narrative about US postwar growth, including the "productivity slowdown" of the 1970's or the 1990's 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 the persistence of the slowdown (Antolin-Diaz et al., 2017).

Panel (b) presents posterior estimates 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. Just prior to the COVID-19 pandemic, output volatility reached a historically low level of less than 1% in annualized terms. The plot also shows that volatility spikes during recessions, in line with the findings of Bloom (2014) and Jurado et al. (2015), so the random walk specification is flexible enough to capture cyclical changes in volatility as well as permanent phenomena. Online Appendix C presents the estimates of the SV in the idiosyncratic components of individual indicators.

3.2 Heterogeneous lead-lag dynamics

In any filtering problem, prediction errors for the observables map into updates of the estimate of the latent states. In our model, updates to the factor are revisions to the estimate of current economic conditions. A poorly specified process for each observable series will lead to large errors, and therefore to erratic updates of the factor. The revisions in the forecasts of individual series can be thought of as impulse response functions (IRFs) of the series to an innovation in the factor process, i.e. a common shock. In a standard DFM, in which no lags of the factor are included in the measurement

equation ($m = 0$), these IRFs are proportional to the dynamics of the common factor.[12] This proportionality is broken by the inclusion of lags of the factor in the measurement equation, i.e. the presence of the lag polynomial $\mathbf{\Lambda}(\mathbf{L})$ of order m in equation (1).[13] This allows the DFM to pick up richer dynamics in the panel of macro indicators.

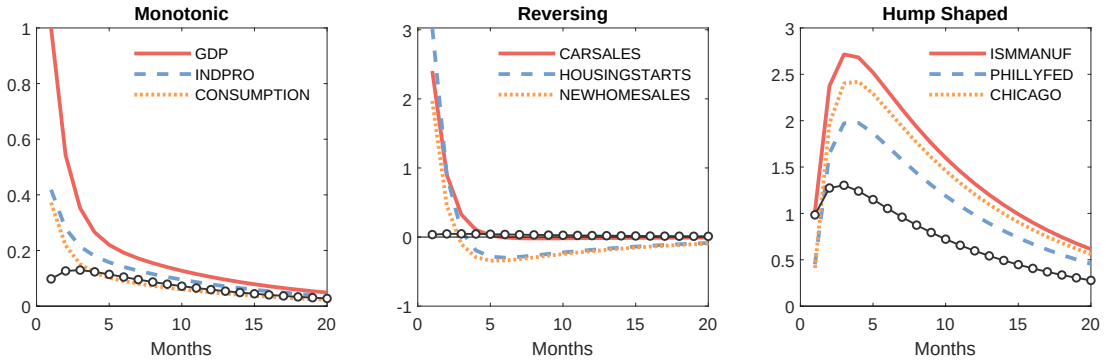
Figure 4 plots these IRFs for different variables and groups them according to their estimated shapes. To better appreciate the differences with a standard DFM, in each panel we superimpose the average IRFs for the same variables from the standard DFM (black line with circular markers). The IRFs reveal that broad aggregates capturing output and production respond with a decaying pattern, where the peak response is on impact but persists for many months (left panel in Figure 4). A number of other variables, particularly those related to investment such as vehicle sales or construction indicators, have a strong initial response which turns negative after a few months before decaying to zero (middle panel). Recall that the variables are expressed in differences, so this indicates an initial overshoot and subsequent decay in levels, consistent with the behavior of pent-up demand for durables (Beraja and Wolf, 2021). Finally, a third group of variables, primarily the business surveys, display clear hump-shaped responses to the common shock (right panel).

Across the three panels, it is visible that the IRFs in the basic DFM without lead-lag dynamics inherit the shape of the IRFs of the survey variables. Moreover, updates to the factor do not lead to large responses in the forecast of the variables in the monotonic group and map into essentially no revisions in the forecasts for the reversing variables. In other words, in the basic DFM business surveys have a disproportionate weight in the computation of the factors, owing to their high contemporaneous correlation with the business cycle and high degree of persistence, and the resulting forecast for the factor will inherit their properties. This is unappealing when data series that mean-

[12] Specifically, the IRF of variable y_i at time t and horizon h is given by that of the factor, scaled by the loading coefficient λ_i: $\frac{\partial y_{i,t+h}}{\partial \varepsilon_t} = \lambda_i \frac{\partial f_{t+h}}{\partial \varepsilon_t}$.

[13] D'Agostino et al. (2016) refer to this case as "dynamic heterogeneity" and explore it in a six-variable model. The larger size of our dataset allows to uncover three broad patterns of impulse responses, illustrated in Figure 4.

Figure 4: IRFS OF SELECTED VARIABLES TO A COMMON SHOCK



Notes. IRFs of different variables to an innovation in the process of the dynamic factor (an innovation to ε in equation (3)). In each panel, the black line with circular markers shows the average across IRFs with homogeneous dynamics ($m = 0$), while the solid red, orange dotted and dashed blue lines show the IRFs for selected variables with heterogeneous dynamics ($m = 1$). The three panels group these IRFs into categories based on their shape ('monotonic', 'reversing' and 'hump-shaped').

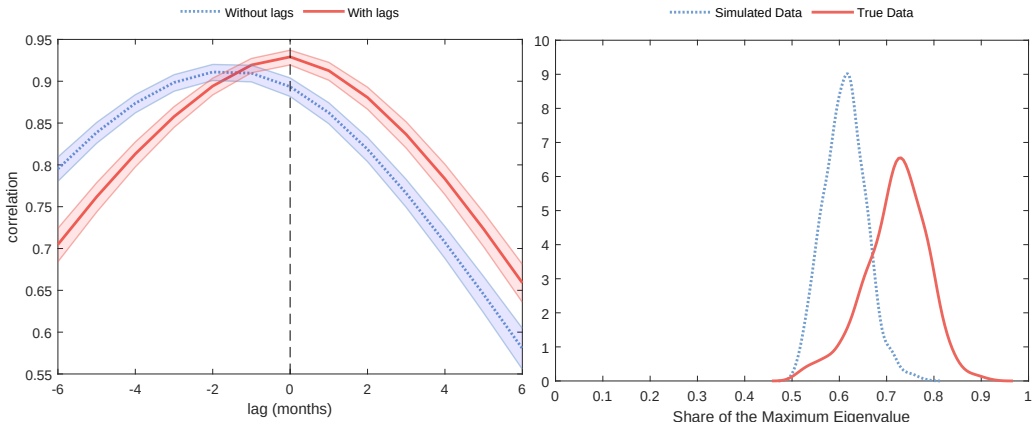
revert more quickly may provide a meaningful signal of economic activity not captured by the surveys. Indeed, in the basic DFM the variables that tend to lead the cycle, such as housing and durable spending, fall into the group of "reversing" variables and therefore have very small loadings. This implies that the standard model perceives them as largely uncorrelated with the common factor and discards useful information in these indicators.

To analyze the lead-lag dynamics further, Figure 5, Panel (a) shows that in a specification that does not include heterogeneous dynamics ($m = 0$), the cross-correlogram of the common activity factor with respect to the implied month-to-month variation of GDP growth is not centered around 0. This means that the estimated factor lags GDP growth, which implies that the nowcasts are inheriting a delay. The same cross-correlogram peaks at lag zero for a model with $m = 1$. As we will show in Section 4, this difference is most visible around turning points of the business cycle, which improves the nowcasting performance of the model.

An implication of our modelling choice of one factor ($k = 1$) and heterogeneous

Figure 5: UNPACKING HETEROGENEOUS DYNAMICS

(a) Cross correlogram **(b)** Static vs. dynamic factors



Notes. Panel (a) presents 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 better captures contemporaneous movements in real GDP growth. Panel (b) is based on rewriting our DFM as a two-factor model and estimating it on our data set. We check how close the estimated factor covariance matrix of this model is to reduced rank. In particular, we compute the share of its largest eigenvalue for each posterior draw (solid red curve).

dynamics ($m > 0$) is that we take the view that a single aggregate innovation is responsible for the bulk of fluctuations in our panel, even though its propagation is different across variables. This parsimonious specification contrasts with the alternative modeling choice, followed e.g. by Stock and Watson (2012) or Bok et al. (2018), of using more factors. The two specifications are related, as 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.[14] Which specification is preferred amounts to asking how close to reduced rank is the covariance matrix of a multiple factor specification in the data. Therefore, we estimate a homogeneous DFM with two static factors for our data set, and compute,

[14] A large part of the literature calls the first specification a "dynamic factor model" and the second a "static factor model", 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. (2016).

for each posterior draw, the share of the largest eigenvalue in the factor covariance matrix. The result, displayed in panel (b) of Figure 5, is a modal share of about 75% of the largest eigenvalue. This is higher than an estimate of the same quantity using simulated data where the restriction is satisfied. Our results support the idea that a single aggregate innovation, transmitted heterogeneously across variables, can capture common fluctuations in real activity variables.[15]

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

In a DFM, the update of the factor estimate in response to the release of new information about the j-th observable, can be understood in terms of the "influence function",

$$E(f_{t_k}|\Omega_2) - E(f_{t_k}|\Omega_1) = w_{j,t} \quad y_{j,t_j} - E(y_{j,t_j}|\Omega_1) \quad , \tag{10}$$

where Ω_2 is an information set containing additional observations relative to information set Ω_1, and $y_{j,t_j} - E(y_{j,t_j}|\Omega_1)$ is the "news", the forecast error based on the old information set.[16] The weight $w_{j,t}$ is the slope of the influence function. In a linear Gaussian state space model this corresponds to the j-th column of the Kalman gain matrix and is invariant to the size of the prediction error. As a consequence, the update of the factor is a linear function of the surprise in the release of the data. On the contrary, in our model $w_{j,t}$ takes the form

$$w_{j,t}(y_{j,t_j}) = \frac{E \quad (f_{t_k} - f_{t_k|\Omega})(f_{t_j} - f_{t_j|\Omega}) \quad \Lambda_j'}{\Lambda_j E \quad (f_{t_j} - f_{t_j|\Omega})(f_{t_j} - f_{t_j|\Omega}) \quad \Lambda_j' + \sigma^2_{\eta_{j,t_j}} + \delta_{j,t_j}\sigma^2_{o,j}}. \tag{11}$$

[15] This innovation can span multiple *structural* macroeconomic shocks. Identifying these structural shocks separately would require additional variables, such as prices, as well as additional assumptions.

[16] When more than one variable is released, equation (10) can be used to obtain a "news decomposition", i.e. the contribution of each variable to the factor update, see Banbura and Modugno (2014). The separate notation for t_j and t_k allows for the time period for which the news about variable j arrive, and the period for which the factor estimate is updated, to differ. In general, the influence function in the robust statistics literature can be thought of describing the effect of an additional observation (in this case the realized prediction errors) on a statistic of interest (the factor's estimate), see Hampel et al. (1986).

The expectation terms $E\ (f_{t_k} - f_{t_k|\Omega})(f_{t_j} - f_{t_j|\Omega})$ in (11) denote the filtering uncertainty about the common factor. In a model with stochastic volatility, an increase in the presence of large errors across many variables will lead to an upward revision in the estimated time-varying volatility of the common factor. The derivative of $w_{j,t}$ with respect to this common volatility is positive, which means that in periods of higher aggregate uncertainty, the signal-to-noise content of all variables increases, and the factor estimates will be more sensitive to incoming news.[17] On the contrary, when the stochastic volatility of the idiosyncratic components, $\sigma^2_{\eta_{j,t_j}}$, increases in the denominator, the signal-to-noise ratio declines and the model becomes less sensitive to news. Equation 11 makes clear that in our model $w_{j,t}$ depends on the data itself. The reason is the term δ_{j,t_j}, which takes the form

$$\delta_{j,t_j} = (((y_{j,t_j} - E(y_{j,t_j}|\Omega))^2/\sigma^2_{o,j} + v_{o,j})/(v_{o,j} + 1), \tag{12}$$

where $v_{o,j}$ is the estimate of the degrees of freedom of the t-distributed component of variable j. δ_{j,t_j} increases with the size of the error, and therefore lowers the weight. As a consequence, large idiosyncratic errors are discounted as outlier observations containing less information. As $v_{o,j} \to \infty$, the outlier component becomes Gaussian and the influence function collapses to the linear form. Thus, the SV makes the influence functions time-varying, and the Student-t component makes them non-monotonic.

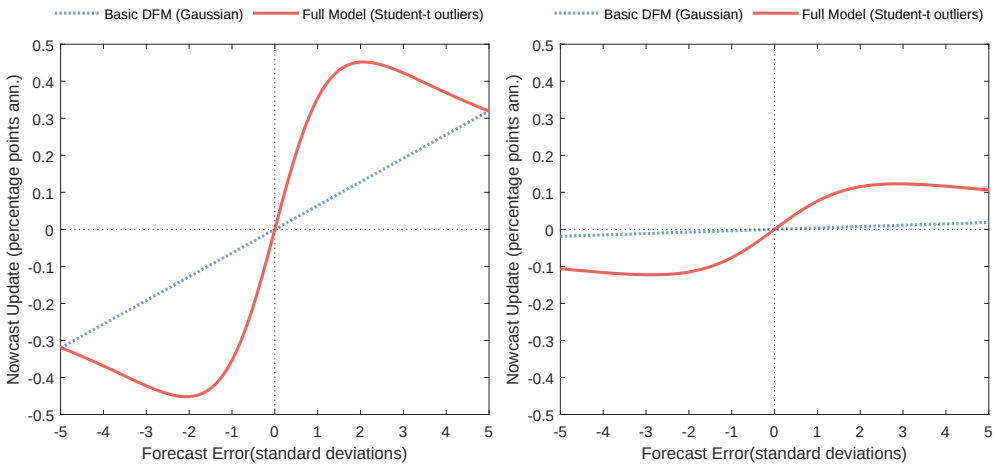
Figure 6 plots the influence functions for two example variables, industrial production and car sales, both for the linear case (without outliers) and the nonlinear case (with outliers).[18] For a model with Gaussian innovations, the update of the factor is a line with slope equal to $w_{j,t}$. These are displayed as the dotted blue lines. As can be seen, noisy variables such as car sales have very low weights, meaning that surprises to

[17] As a consequence, it is possible for the factor uncertainty to increase as a consequence of the release of new data. 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).

[18] The influence functions for all variables are presented in Online Appendix C.

Figure 6: INFLUENCE FUNCTIONS

(a) Industrial production **(b)** Light weight vehicle sales

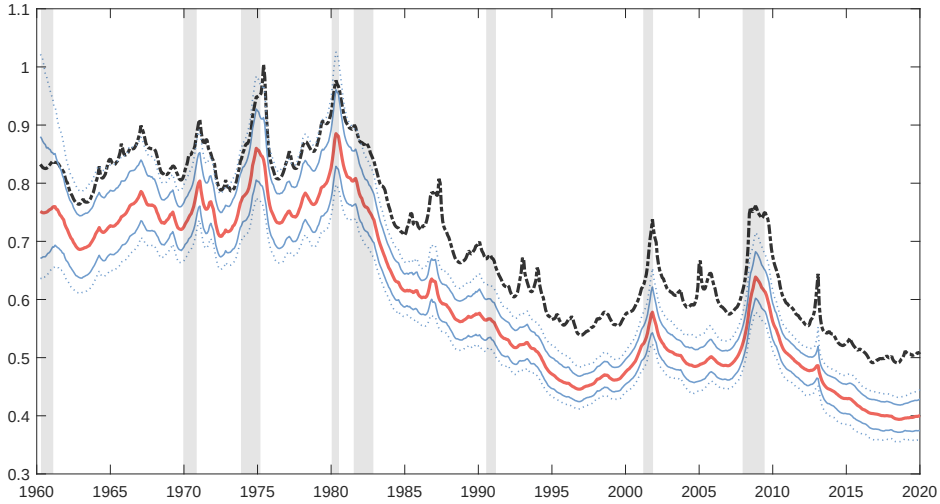


Notes. Influence functions for industrial production and light weight vehicle sales. An influence function indicates by how much the estimate of the common factor is updated when the release in the variable is different from its forecast and thus contains "news." The dotted blue line plot these influence functions in the Gaussian case, while the red lines represent the Student-t case. As shown in equations (11)-(12) the model with fat tails allows these functions to be nonlinear and nonmonotonic. As the influence function is in general time-varying, we report the average influence function in a representative year. The influence functions for other variables are provided in Online Appendix C.

this variable lead to small updates to the current factor. With t-distributed outliers, the influence functions are now S-shaped, shown as the red solid lines of Figure 6. 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 was manually replaced with a missing observation.

The modeling of outliers and heterogeneous lead-lag dynamics 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

Figure 7: UNCERTAINTY INDEX WITH AND WITHOUT OUTLIER TREATMENT



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. (2015). This index is computed as shown in equation (13) 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.

implies an increase in the slope of their influence functions. 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.

The modeling of outliers also interacts directly with the presence of SV. To illustrate this, we compute an index of uncertainty following Jurado et al. (2015) using our model.[19] Figure 7 reports the mean and HPD bands of the uncertainty index

$$\mathcal{U}_t \equiv \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{\lambda_i^2 \sigma_{\varepsilon,t}^2 + \sigma_{u,i,t}^2}{\hat{s}_i^2}} \tag{13}$$

where $\hat{s}_i$ is a variable-specific scaling factor capturing differences in average standard

[19] Online Appendix C presents the estimated SV of all idiosyncratic components of the model.

deviation across variables. The advantage of $\mathcal{U}_t$ is that, apart from reflecting the time variation in the volatility of the common factor, $\sigma_{\varepsilon,t}$, as in Figure 3 (b), it captures 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, in contrast to the one of Jurado et al. (2015) 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 fewer transitory 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 7 displays the same index calculated from a version of the model which does not feature the outlier component. This index is above our estimate throughout the sample, since the presence of outliers in the data, if not explicitly modeled, inflates the estimate of the volatility of the idiosyncratic component. Importantly, the broken black line exhibits more frequent and larger spikes. Some of them, like the one visible in 2005, can be attributed to short-lived natural disasters such as hurricane Katrina. Our results suggest caution when interpreting economic uncertainty indices in the presence of fat-tailed innovations that may not wash out in the aggregate.[20]

4 Real-time out-of-sample model evaluation 2000-2019

This section constructs daily estimates of US GDP growth, and formally evaluates the performance of our model relative to benchmark competitors, both in terms of point and density forecasting. To mimic the exercise of a forecaster who updates her information set in real time, we build a data base of unrevised vintages of data for

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

each point in time, every day between January 2000 to December 2019. This involves carefully addressing various intricacies of the data vintages, such as methodological changes that occur over time. Furthermore, re-estimating the model every day that the information set is updated over 20 years is computationally very intensive. It is made feasible thanks to our efficient algorithm and by exploiting modern cloud computing infrastructure. Details are provided in Online Appendix D.

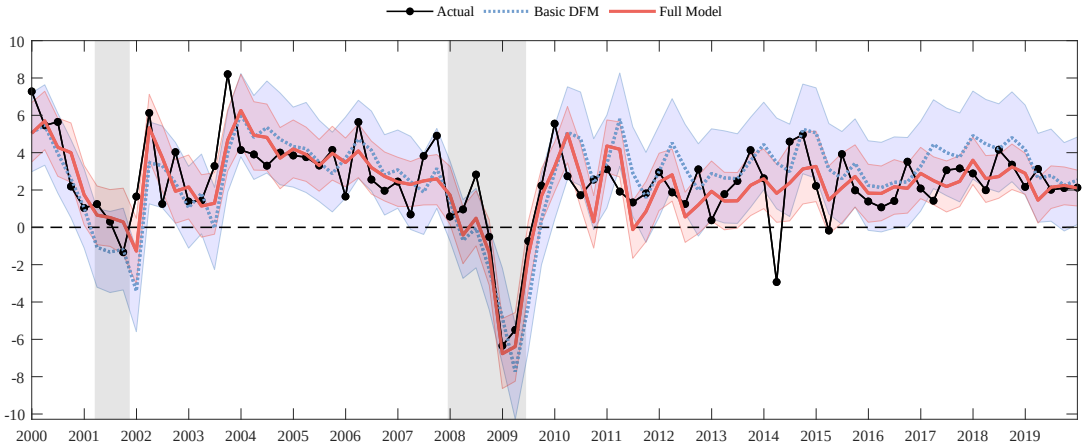
4.1 Model forecasts and GDP releases over time

Figure 2 presents our daily estimate of current-quarter real GDP growth, together with its estimated volatility. This time series represents a daily snapshot of current economic conditions produced by the model. In the Online Appendix we report additional outputs from the model, including daily estimates of the probability of recession as well as a measure of growth-at-risk in the spirit of Adrian et al. (2019).

Figure 8 turns to a formal comparison of the model predictions with the official GDP data subsequently published by the BEA. It plots the predictions of two versions of the model: our full Bayesian DFM (shown as the red solid line), as well as a basic DFM that is estimated on the same data set but does not feature time-varying trends, SV, heterogeneous dynamics or fat tails, such as the model of Banbura et al. (2013) (dotted blue line). The forecasts are produced about 30 days after the end of the reference quarter, just before a first ("advance") estimate of GDP growth is published by the BEA. We compare the model forecasts with the "final" or third release of GDP published by the BEA two months later, which is plotted with black circles. The figure shows that both models track the broad contour of fluctuations in real GDP, including the recessions and recoveries in 2001 and 2008-09. Importantly, however, the basic model displays very visible drawbacks. A persistent upward bias is evident after around 2010, reflecting a failure to capture the decline in long-run growth which materialized in this decade (see also Antolin-Diaz et al., 2017). Furthermore, the comparison of the two

models in real time highlights the advantage of modeling heterogeneous dynamics. The full model better captures both the magnitude of the contractions and the timing of the recoveries. This is because the model can balance the information in indicators of durable consumption and investment, which serve as leading indicators of the recovery, with that of the more sluggish business surveys.

Figure 8: MODEL NOWCASTS AND PUBLISHED GDP GROWTH COMPARED



Notes. The black line represents the third release of real GDP growth in the United States as published by the BEA. 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.

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%.[21] 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

[21] Applying a formal test for the correct calibration of the density forecast (Rossi and Sekhposyan, 2019) rejects (at 10% level) the correct specification for the basic DFM model, but does not reject when applied to the forecasts of the full model. This is true when the test is applied to the entire density, as well as when this is applied to the left and right side of the density separately.

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 Real-time evaluation and comparison to alternative models

Table 1 provides a formal forecast evaluation of our model against several alternatives. We again compare our full model against a basic DFM. We also add a simple AR(1) model for quarterly GDP growth data, as well as the DFM maintained by the New York Fed staff, described in Bok et al. (2018).[22] The table formally evaluates the point and density forecast accuracy relative to the third release of GDP.[23]

Panel (a) focuses on point forecasts, by comparing the root mean squared error (RMSE) across models. In each column, the different lines show the RMSE of a given model for different forecast horizons. A more accurate forecast implies a lower RMSE. Recall that GDP is quarterly (covers a period of 90 days) and is first released 30 days after the reference quarter. Predictions produced 180 to 90 days before the end of a given quarter are *forecast* of the next quarter; forecasts at a 90-0 day horizon are *nowcasts* of the current quarter, and the forecasts produced 0-30 days after the end of the quarter are *backcasts* of the last quarter. In brackets we report the p-value from the Diebold and Mariano (1995) test of the null hypothesis that a given model performs as well as our full Bayesian DFM, against the alternative that the full model performs better. Reading the table from top to bottom, it is visible that models become more accurate as the forecasting horizon shrinks and more information comes in. Importantly, our

[22] The New York Fed Staff Nowcast is updated weekly. Historical nowcasts are available online, which allows a direct comparison. See https://www.newyorkfed.org/research/policy/nowcast.html.

[23] As there are three major releases and GDP gets revised over time, an important question is which vintage of GDP is taken as the "ground truth" against which forecasts are evaluated. We focus on the third ("final") release, as the majority of revisions occur in the earlier releases. We have explored evaluating the forecasts against earlier 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. The results are available upon request. If the objective is to improve the performance of the model relative to the first official release, then an explicit model of the revision process would be desirable.

Table 1: OUT-OF-SAMPLE PERFORMANCE OF DIFFERENT MODELS

(a) Point forecasting: RMSE	**Full Model**	**AR(1)**	**Basic DFM**	**NY Fed**
–180 days	2.25	2.36	2.53	–
		[0.36]	[0.02]	–
–90 days (start reference quarter)	2.14	2.33	2.57	2.36
		[0.28]	[0.02]	[0.29]
–60 days	1.88	2.14	2.21	2.18
		[0.23]	[0.01]	[0.05]
–45 days (middle reference quarter)	1.65	2.14	2.09	1.98
		[0.23]	[0.01]	[0.02]
–30 days	1.66	2.11	2.09	1.83
		[0.10]	[0.01]	[0.07]
0 days (end reference quarter)	1.48	2.12	1.98	1.67
		[0.04]	[0.00]	[0.08]
+30 days (first release)	1.48	2.14	1.96	1.60
		[0.03]	[0.00]	[0.13]
(b) Density Forecasting: Log Score	**Full Model**	**AR(1)**	**Basic DFM**	**NY Fed**
–180 days	–2.16	–2.42	–2.40	–
		[0.00]	[0.36]	–
–90 days (start reference quarter)	–2.09	–2.40	–2.37	–2.28
		[0.00]	[0.33]	[0.05]
–60 days	–1.98	–2.34	–2.24	–2.17
		[0.00]	[0.09]	[0.01]
–45 days (middle reference quarter)	–1.89	–2.35	–2.18	–2.08
		[0.00]	[0.09]	[0.00]
–30 days	–1.88	–2.34	–2.18	–1.99
		[0.00]	[0.02]	[0.00]
0 days (end reference quarter)	–1.81	–2.34	–2.14	–1.91
		[0.00]	[0.00]	[0.00]
+30 days (first release)	–1.80	–2.35	–2.13	–1.87
		[0.00]	[0.00]	[0.00]

Notes. Comparison of the forecasting performance of different models: the full Bayesian DFM; an AR(1) model on quarterly GDP data; the basic DFM; the New York Fed Staff Nowcast. We report the performance in RMSE (top panel) and Log score (bottom panel) across various forecasting horizons. For the first three column the sample is 2000-2019. For the last column, it is 2002-2019. For each of the alternative models we report the p-value associated with the null hypothesis that a given model performs as well as the full Bayesian DFM model against the alternative that the full model performs better. The test is computed using the Diebold and Mariano (1995)'s statistic with small-sample correction as suggested by Clark and McCracken (2013). Note that the New York Fed's model is a frequentist model that does not produce density forecasts. We construct the associated density forecasts by resampling from past forecast errors as in Bok et al. (2018).

full Bayesian DFM produces the lowest forecast errors at all horizons. It is significantly better than the basic DFM at all horizons, and beats the AR(1) model from a 30-day horizon onwards at conventional significance levels.[24] Our model also outperforms the NY Fed Model with economically large improvements. At the 45-day horizon, the RMSE from our model is 17% lower and highly statistically significant. The differences decline by the time of the first release of GDP, 30 days after the reference quarter.

Panel (b) turns to comparing the Log score, a common metric for density forecast evaluation, across the same models and horizons as the previous panel. A larger average Log score indicates a more accurate density forecast. Being estimated with frequentist methods, the New York Fed's model does not produce density forecasts. We construct the associated density forecasts by resampling from past forecast errors as in Bok et al. (2018). In brackets we again report the p-value from the Diebold and Mariano (1995) test. It is evident from panel (b) that the relative density forecasting performance of our Bayesian DFM is even stronger than the point forecasting performance. When considering conventional significance levels, the full model beats all alternatives at all horizons, except for the basic DFM, which is not significantly worse and very far horizons of 180 and 90 days ahead of the end of the reference quarter. Overall, the table highlights the strong performance of the Bayesian DFM at producing point and density forecasts for US real GDP growth.

4.3 Comparison to Federal Reserve and survey expectations

In addition to assessing the performance of our Bayesian DFM relative to other formal econometric models, we compare its forecasts to those of the Survey of Professional Forecasters (SPF) and its individual participants, as well as to the Federal Reserve Staff's forecasts produced for the Federal Open Market Committee (FOMC). A comparison

[24] The relatively good performance of the AR(1) is somewhat misleading. As we show in Online Appendix E, its low RMSEs are driven by the fact that the AR(1) does very well on average in quiet times, but completely misses periods of crisis. This is important to consider when comparing the significance of the basic DFM and the AR(1) relative to each other, which we do in the same Online Appendix.

against professional and institutional forecasts is generally considered a very high bar for statistical models, as they have access to a large information set including news, political developments, and information about structural change (see, e.g., Sims, 2002). Consensus measures such as the SPF median are an even higher benchmark as they aggregate the opinions of a multitude of forecasters.

Panel (a) of Figure 9 compares the RMSE for the GDP nowcast of our Bayesian DFM against the RMSEs of the current-quarter GDP forecasts of individual participants in the SPF, and the associated SPF survey median. We also include the NY Fed Staff Nowcast considered in the previous section. These forecasts are evaluated in the middle of the quarter, so correspond to horizon -45 in the evaluation above. The chart provides this information as a scatter plot, given that the comparison of the individual nowcasts against ours are carried out over different samples. This is due to the fact that SPF panelists drop in an out of the survey, so we compare each for the overlapping set of observations.[25] Observations above the 45-degree correspond to forecasts that are less accurate than the ones obtained from our model. The chart delivers several insights. First, most individual SPF forecasters produce larger error than our model. Our model outperforms 80% of all individual forecasters. Second, our forecasting performance is very similar to, and statistically indistinguishable from, the SPF median. Third, in line with the previous section, the Bayesian DFM outperforms the NY Fed nowcasts.

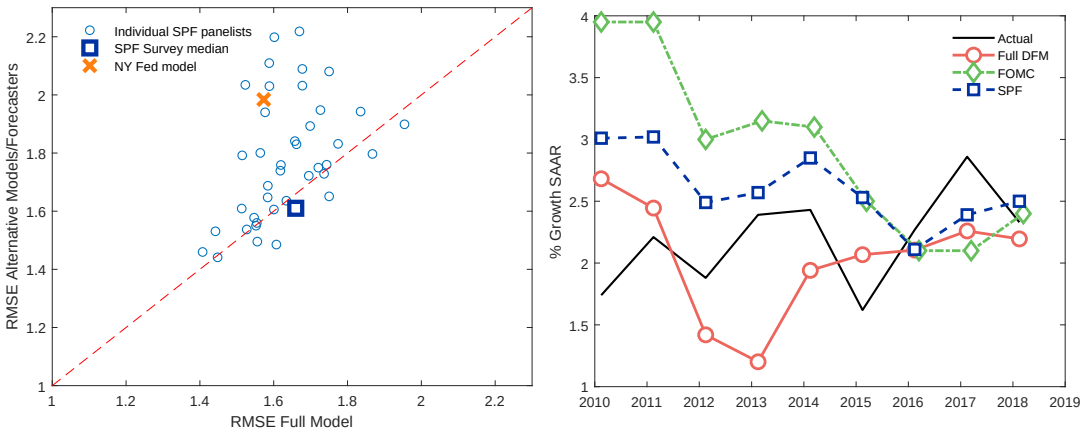
We also compare the quarterly GDP nowcasts of our model against the Federal Reserve Staff Projections for GDP included in the Greenbook (GB, nowadays known as "Tealbook"). This comparison is not feasible at a fixed 45-day horizon, as the GB is prepared prior to FOMC meetings, which occur eight times per year, so the exact horizon varies every meeting. Broadly speaking, there is a meeting early in the quarter (the median happening 66 days before the end of the quarter) and another one later in the quarter (18 days). By matching the exact information set with the date of the

[25] Appendix E explains how we treat individual forecasts, given that the panel of participants is unbalanced.

Figure 9: MODEL PERFORMANCE RELATIVE TO SPF AND FED FORECASTS

(a) Quarterly GDP nowcasts

(b) 2-year ahead GDP forecasts



Notes. Panel (a) presents a scatter plot of the root mean squared error (RMSE) for real GDP nowcasts of our full Bayesian DFM against the RMSE of alternative forecasting models: individual forecasts from the Survey of Professional Forecasters (SPF); the median of the individual SPF forecasts; the New York Fed Staff Nowcast. Points above the 45-degree line indicate that our model is more accurate. Panel (b) instead provides a comparison for the 2-year ahead GDP forecasts against the actual realization at different points in time. This is shown for our model; the SPF; and forecasts produced by the FOMC.

GB, we find that our model's forecasts corresponding to the early meeting are not significantly different in terms of RMSE to the GB's, whereas the GB dominates by the time of the late meeting.[26] Thus, relative to the results of Faust and Wright (2009), we find that a carefully specified model using many indicators of real activity can come close to the performance of the Greenbook.[27] Yet, in line with earlier findings of Romer and Romer (2000), we document that Federal Reserve Staff forecasts continue to be a tough benchmark for formal models, appearing to possess considerable additional information (Nakamura and Steinsson, 2018).

Panel (b) considers longer-horizon forecasts, and plots 2-year ahead GDP growth

[26] In the early meeting, our model has a RMSE of 1.99 vs. 1.88 of the GB, p-value of 0.23, whereas for the late meeting, the RMSE of our model 1.37 vs. 1.67 of the GB $p - value = 0$

[27] Faust and Wright (2009) report that "large data methods" produce losses 30% higher than the GB for the 1984-2000 sample, and even higher in the 1979-2000 sample, when there is a recession in the sample. Moreover, they report that these models are clearly outperformed by simple AR models, which is not true in our case.

forecasts made at different points in time against the actual realization of GDP growth. We provide a comparison of the 2-year ahead forecasts from our Bayesian DFM relative to the SPF survey median and the "Summary of Economic Projections" published by the Federal Open Market Committee since 2008. The out-turns for the 2-year ahead forecasts thus start in 2010. The chart shows that the SPF median and FOMC forecasts overestimate GDP in the first half of the 2010's, while our model moves more symmetrically around the actual outcome. While our model was not built with the purpose of predicting longer horizons, the slow-moving growth component avoids excessive mean reversion in short-run predictions and eliminates the upward bias in longer-horizon forecasts. Finally, towards the later part of the sample, the different forecasts converge to very similar magnitudes.[28]

4.4 Additional evaluation results

We have examined the out-of-sample performance of our model across many more dimensions, and present a host of additional evaluation results in Online Appendix E. This includes showing the RMSE and Log score from Table 1 over a continuous forecast horizon (rather than at a selected number of horizons), reporting the RMSE and Log score through time (rather than across horizons), using alternative evaluation metrics (mean absolute error and continuous rank probability score), and evaluating the individual contribution of different model components (trends and SV, heterogeneous dynamics, fat tails) to the forecasting performance. Finally, the Online Appendix also demonstrates how our model can be used to asses tail risks in real time.[29]

[28] Over the 2010-2018 period, the RMSE of our model is 0.62 percentage points compared to 0.66 of the SPF and 1.14 of the FOMC. The average forecast error, a measure of bias, is 0.37 for our model compared to 0.58 of the SPF and 1.04 of the FOMC.

[29] See also Carriero et al. (2020) for an alternative approach to forecasting tail risks.

5 Understanding macro data during and after COVID-19

The COVID-19 pandemic and associated recession pose a challenge for macroeconometric models.[30] First, the *scale* of the drop in economic activity is unseen in postwar history; second, the *speed* at which events unfolded poses challenges for nowcasting models that rely on traditional monthly economic indicators; and third, the sectoral *composition* of the shock is rather unusual, hitting particularly hard sectors such as services which usually display little business cycle fluctuations, and breaking the typical comovement patterns across macroeconomic variables.[31] The framework that we propose in this paper proves to be critical both to track in real time and to interpret the developments in macroeconomic time series observed during 2020. In fact, attempts to estimate a basic linear and Gaussian DFM lead to unstable estimates and, often, failure of the MCMC algorithm to converge unless one discards (or winsorizes) the observations surrounding the lockdown. Below we describe how the combination of new components enable us to understand the data during this episode. Furthermore, as a separate contribution, we provide a methodology to incorporate into our framework a number of recently developed "alternative" high-frequency indicators which offer a potentially more timely but partial reading of economic activity.

5.1 The Great Lockdown through the lens of the model

Figure 2 previewed the daily reading of the data through the lens of the model during the period between March and August 2020. The main feature that stands out is not just the large drop in underlying economic activity, but also the large increase in the volatility of the common factor. Therefore, in real time the model conveys the idea that

[30] The specification choices for the model and the analysis preceding this section was completed before the end 2019, when initial drafts of this paper were circulated. None of the priors, settings, or modeling choices were altered ex-post to better fit the 2020 data, so the observations for 2020 constitute an actual out-of-sample set for our model.

[31] 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 and Song (2020), and Cimadomo et al. (2020).

not only the outlook is worsening, but uncertainty around it is increasing.

Against the backdrop of this large increase in aggregate volatility, the COVID-19 episode is unique because of the important changes in the sectoral behaviour of macroeconomic time series. Over the typical business cycle, output, income, investment, consumption and employment move closely together, with investment typically more volatile than consumption, and hours worked and other measures of labor utilization moving as much as output. Instead, the COVID-19 recession has led to a large contraction in sectors that are usually not very cyclical and therefore have a small loading on the factor, such as services consumption. In addition, heterogeneity in the impact of the recession across households, as well as aggressive policy interventions to support income and expand unemployment benefits, have decoupled series that comove, such as filings for unemployment claims and aggregate consumption which are usually strongly negatively correlated.

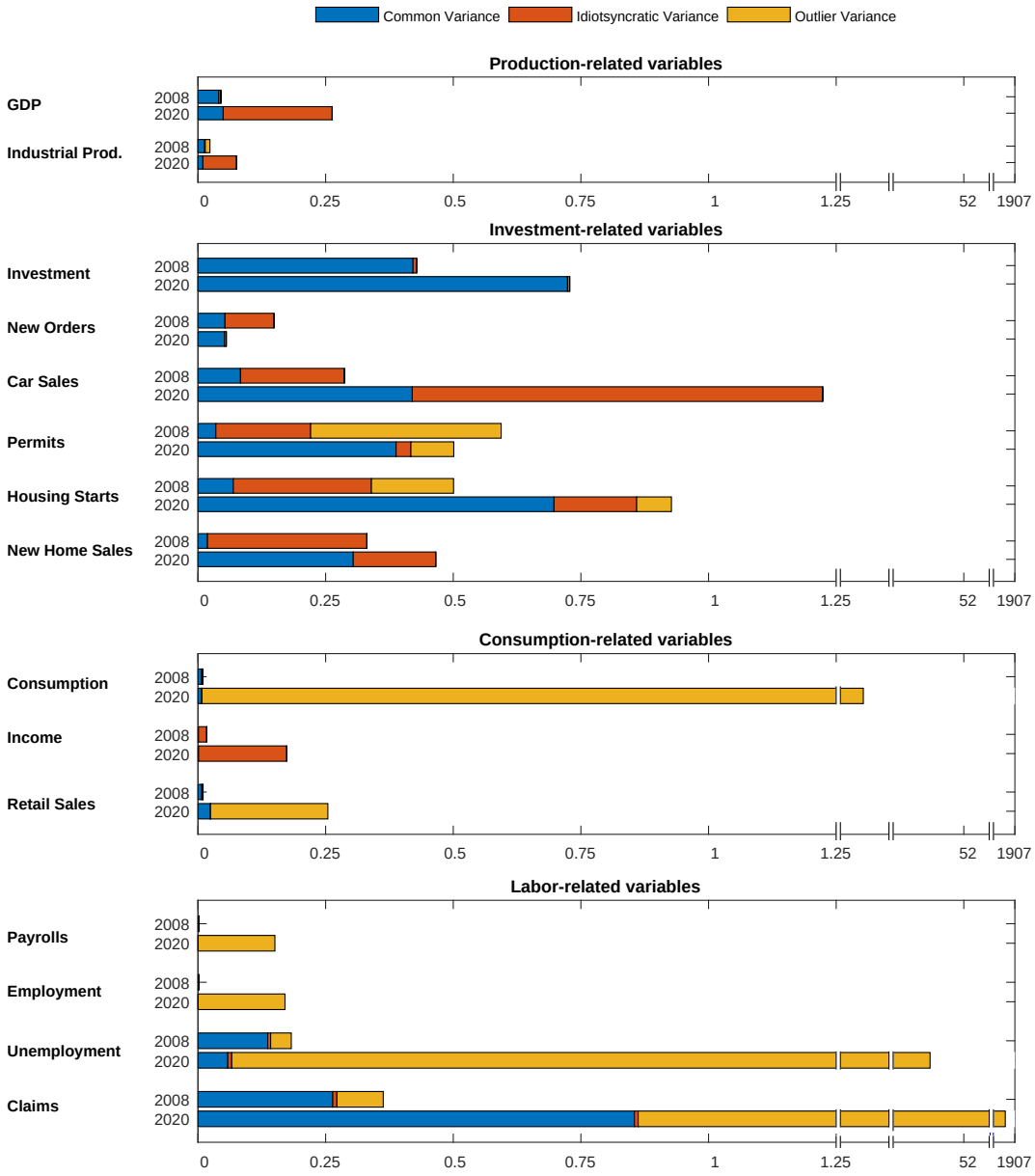
In a DFM the average comovement pattern is reflected in the estimated factor loadings $\mathbf{\Lambda}(L)$. In our model, variables can deviate from this pattern in a persistent way, through the time variation in the volatilities of the idiosyncratic components, or in a transitory way, through the Student-t component.[32] This allows the model to track business cycle fluctuations as a combination of changes in aggregate activity and overall uncertainty, and a mix of transitory and persistent changes in volatility of individual indicators that is unique in each downturn. A variance decomposition of the variables for particular episodes summarizes this. Recall from equation (1) that each series can be decomposed into the sum of the different components. Their innovations are independent, so the variance of each variable at each point in time can be expressed as the sum of the variances of each of its components.[33]

Figure 10 reports such a variance decomposition for selected variables during

[32]This means that which variables are important for the estimation of the factor varies over time even if the loadings are fixed.

[33]This is not true for standard deviations, because the square root of a sum is not the sum of the square roots. By using variances instead of standard deviations, the scale of the figure does not have easily interpretable units.

Figure 10: VARIANCE DECOMPOSITIONS: 2008 VS. 2020



Notes. This figure decomposes the monthly realized variance of individual variables into its different components: variance of the common factor (dark blue); variance of the idiosyncratic component (red); variance of the Student-t component (yellow). This is calculated separately for the years 2008 and 2020. Variables are grouped into different categories, separately showing production-related, investment-related, consumption-related and labor-related indicators.

the period January-August 2020. We group the indicators into conceptually related categories: production, investment, consumption and labor. For comparison with the Great Recession, we do the same exercise for the period January-December 2008. The reader should note that the increase in the variance for some variables in 2020 is massive, so that the horizontal axis displays several scale breaks. For instance, the increase in initial claims in April 2020 was 332 times the standard deviation of the previous 35 years. As a consequence, the increase in variance is even bigger. The blue bars capture the part of the variance that is attributed to the common factor. For most of the variables, this segment is larger in 2020 than in 2008. The COVID-19 recession thus appears to be a larger version of the standard business cycle. This is particularly true for investment-related variables, such as new orders of capital goods, construction indicators, and car sales. In line with the usual low cyclicality of consumption, the blue bars are comparatively small for consumption-related indicators in both episodes. The red bars in turn capture the part of the variance which is idiosyncratic to each series, and the yellow bars capture the outlier component, which is also idiosyncratic but transitory in nature. The striking fact about the COVID-19 episode is the presence of massive outliers *only* on consumption and labor-market related indicators, precisely those series that are most directly affected by the public health interventions intended to curb the spread of the virus. Interestingly, we observe a larger than usual contribution of the idiosyncratic and outlier components for housing variables during the 2008 recession, possibly reflecting the nature of the downturn, which had exceptional swings in mortgage and housing markets at its core. The fact that fluctuations in investment variables are largely explained by the common factor in both episodes leads us to speculate that the common factor in this class of models mostly captures the transmission of aggregate shocks via fluctuations in investment, independently of the exact nature of the shock that triggered the recession.

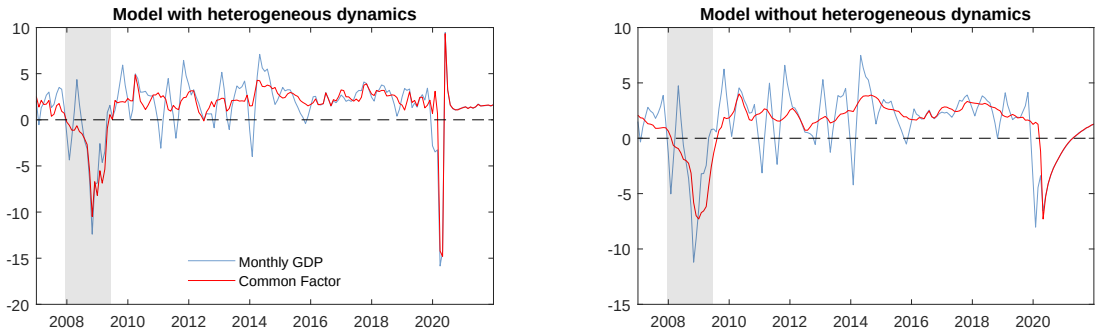
These patterns of changing relative variances would be impossible to capture in

models without SV and fat tailed outliers. Figure 11 examines additional ways in which these components affect the behavior of the model and the resulting interpretation of the data. Panel (a) focuses on the heterogeneous dynamics. We report the estimation results using the data as available on June 1, 2020 for two versions of the model. In the left panel, the full model is used, whereas in the right panel, the heterogeneous dynamics are switched off (i.e., $m = 0$). In both cases, we report the estimated posterior mean of (latent) monthly GDP growth, the common component, and their forecasts through 2021. As discussed in Section 3.2, in standard DFMs with homogeneous dynamics the responses of the variables to a common shock are all proportional, and the business surveys typically dominate the estimation of the common factor. As a consequence, both the factor itself and the forecasts of all variables inherit the dynamics of survey variables. The heterogeneous dynamics ($m > 0$) break that restriction, allowing hard variables such as GDP growth to display less persistent responses to the same shock. As of June 1, survey data available for May already displayed a partial recovery, with hard data for Q2 still mostly not available. For the model with heterogeneous dynamics (left panel) this information translates into a forecast of a sudden large drop in activity followed by a quick, albeit partial, *rebound* in GDP. For the model with homogeneous dynamics (right panel) the forecast is instead for a shallower but more prolonged recession, similar to the one experienced during the 2008-2009 financial crisis. Ultimately, the "partial V" predicted by the full model proved to be a more accurate description of the second half of 2020. This highlights that the model can capture different shapes of recoveries, in line with the differential behavior of durable and non-durable spending recently highlighted by Beraja and Wolf (2021).

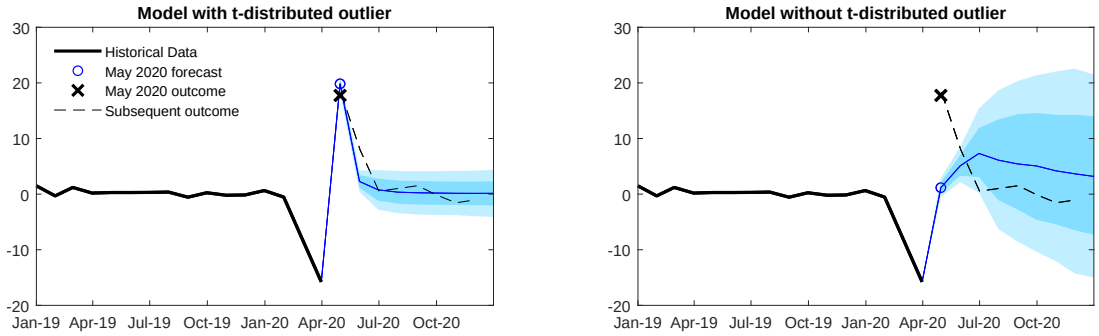
Panel (b) of Figure 11 illustrates the role of the t-distributed outliers. Retail sales in particular experienced an unprecedented 20% drop in April 2020. The left panel reports the full model's forecast for the May 2020 retail sales release, using the data vintage available on June 15, the day before its publication. It is worth noting that

Figure 11: IMPACT OF HETEROGENEOUS DYNAMICS AND FAT TAILS

(a) Projections of GDP growth with and without heterogeneous dynamics (June 1, 2020)



(b) Retail sales forecasts with and without fat tails (June 15, 2020)



Notes. As a case study of the mechanics of our model components in the forecasting process, this figure zooms in on different objects at specific points in time. Panel (a) shows monthly GDP growth and the common factor up to June 1 2020, together with their projection thereafter. It compares this for two model versions, with and without heterogeneous dynamics. Panel (b) focuses on the June 15 vintage of retail sales, which refers to the period of May, and compares the retail sales predictions in model versions with and without the idiosyncratic Student-t component.

the retail sales series displays large, transitory outliers throughout its history, most notably around tax changes in October 1986 and around the 9/11 terrorist attacks (see Figure 1). Therefore, the model interprets a large fraction of the April 2020 drop as one such outlier, and forecasts a strong rebound in May. Moreover, consistent with the transitory nature of the drop and rebound, the idiosyncratic stochastic volatility remains at normal levels for the subsequent forecast horizon. The right panel reports the same forecast for a model in which the t-distributed outliers have been switched

off. As we discussed above, the drop in consumer-related variables such as retail sales was orders of magnitude greater than what one would have expected given the movements in investment-related variables. The only way that a model without fat-tailed outliers can interpret this information is via a massive and persistent increase in the idiosyncratic SV of retail sales. This can be seen in the widening fan chart of the right panel. With no reason to expect a large idiosyncratic outlier of opposite sign, the median path is consistent with the aggregate rebound of the common factor, as discussed in panel (a). Ultimately this forecast turned out to be a worse description of developments in retail sales.

Figures 2, 10, and 11 highlight the main conclusion of examining the data observations for 2020 through the lens of our model. The COVID-19 recession appears, first, as a large drop in aggregate economic activity followed by a quick but partial rebound, accompanied by a persistent increase in the volatility of the common factor; second, it manifests itself by the presence of large, but transitory, swings in those macroeconomic time series capturing the consumer sectors most directly exposed to the lockdown of March-April 2020.

5.2 Incorporating newly available high-frequency data

An additional challenge posed by the COVID-19 recession for tracking economic developments in real time derives from the *timing* of the shock, which had its maximum effect after the declaration of a National Emergency on March 13, 2020. The release calendar for macroeconomic data is such that the most timely indicators are released around the turn of each month, referring to conditions prevailing on average in the previous month. Therefore, the downturn was not fully reflected by standard indicators until their release at the end of April or beginning of May, making nowcasting models potentially less useful at a time where an assessment is urgently needed. In response to this lack of data, a number of novel indicators have been proposed. These include

measures of consumption derived from credit card data, as in Chetty et al. (2020); employment data from the online scheduling provider Homebase and the Real Time Population Survey described in Bick and Blandin (2020); data on restaurant reservations from the mobile application OpenTable; and mobility data provided by Apple.

These "alternative data" provide a potentially more timely reading of economic conditions than traditional indicators, and have therefore received close attention in media and policy circles during the pandemic. Each of these new series however provides only a partial and noisy signal of economic conditions. We propose a method to incorporate them into the DFM framework, in order to exploit their *joint* information in a systematic way, and quantify their contribution to our assessment of the economy. The key difficulty of this task is that these series have an extremely short history, all starting in 2020 and some appearing only as late as April. Estimation of the parameters of the measurement equation (1) thus becomes challenging. Recall that the factor loadings, $\mathbf{\Lambda}(L)$ capture the average responsiveness of each series to the business cycle. For reliable estimation, one would need the series to be available for one or more business cycles. Whether the series is available daily, weekly, or monthly is less relevant.[34] Our key insight is that many of these series are quite closely related to some of the traditional indicators already included in the DFM. For instance, the credit card data from Chetty et al. (2020) explicitly tries to track aggregate consumption. In this case, one would expect the traditional series and the high-frequency proxy to have both similar loadings and correlated idiosyncratic components. This information would be lost if one just appended the new series to the panel.

Our proposal is to qualitatively match, based on the underlying economic concept, each high-frequency series with a traditional monthly indicator, and use the former as conditioning information for the unpublished values of the latter. This distinguishes our approach from that of Lewis et al. (2020) who create an index of economic activity based

[34] Attempting to estimate the model at daily frequency would therefore not help in this dimension.

only on weekly indicators that have been available at least since the 2008 recession. Table 2 lists the novel data series together with the traditional series that we match to each of them. We incorporate the credit card spending series from Chetty et al. (2020) to condition the value of real consumption growth from the BLS; the Homebase employment data and the Dallas Fed Real Time Population survey to match series from the establishment and household surveys, respectively; 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 the number of bookings for seated dinners at restaurants registered by the mobile application OpenTable, which we match to the real retail sales series for food services and drinking places. In addition, we incorporate two series which have been available for a longer sample, the weekly initial claims data from the BLS; and the daily survey of consumers from Rasmussen, which is available since 2004, and which we can link to the two monthly consumer surveys in our panel.[35] In all cases, we verify that the definition of the series (e.g. growth with respect to a year ago) and seasonal adjustment is consistent.

We first aggregate the daily series to monthly frequency. When partial information is available for a given month, we average the observations available within the month.[36] This leaves us with monthly time series that are sometimes as short as one or two observations. We then estimate a linear regression linking the traditional series with the novel proxy. Given the extremely short sample, a Bayesian prior on the coefficients of this regression is employed, tightly centering the intercept around zero and the slope around one. We then use the fitted value for the last observation, when the novel series is available but the traditional series is not, to "nowcast" the value of the latter. In the

[35] The 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. We introduce them into the model assuming a one day publication lag as soon as at least two months of data are available.

[36] Knotek and Zaman (2019) have shown this to be preferable to taking the model to a higher frequency.

Table 2: TRADITIONAL SERIES AND MATCHED HIGH-FREQUENCY PROXIES

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

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. We then matched the series in the DFM, as described in the main text.

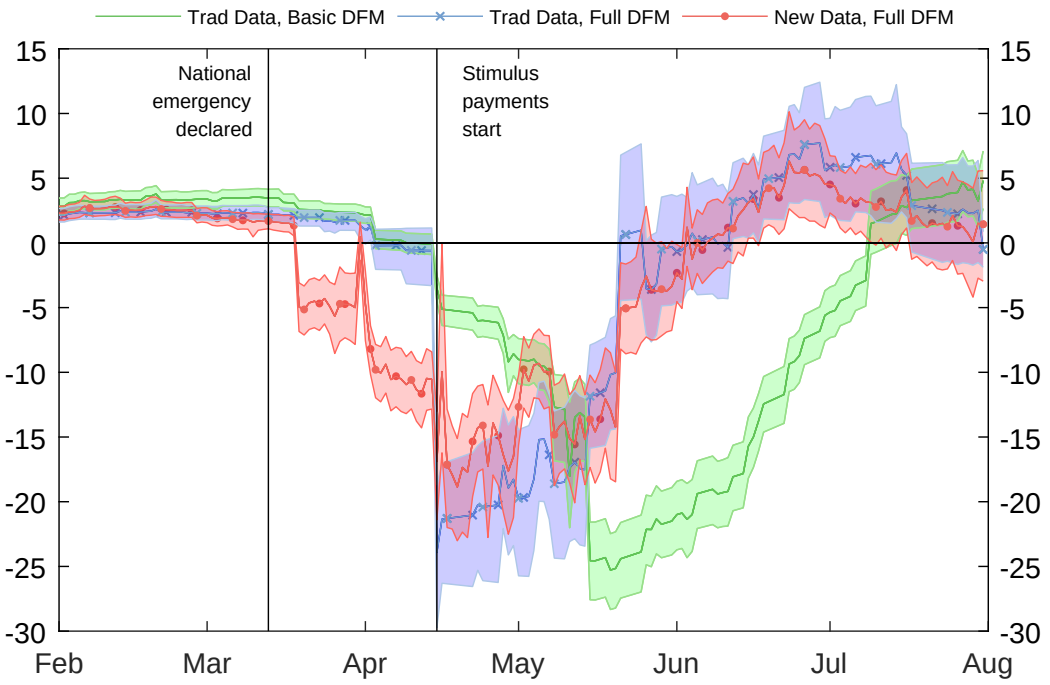
first few months, the prior information will dominate, effectively imposing a close to one-to-one relation between the traditional series and its proxy. Of course, the true relationship between the indicators may not be one-to-one, so the relevant empirical question is whether the extra timeliness added 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. Nevertheless, as data accumulates, the posterior will converge to the true relationship between the two series.[37]

Figure 12 provides a joint assessment of the usefulness of the high-frequency data and of our modeling innovations during the lockdown and subsequent recovery of 2020. We present the daily estimate of US real GDP growth for different combinations of models and data: a basic DFM (solid green) and our full Bayesian DFM (blue with crossed markets) estimated on the traditional data, as well as our full Bayesian DFM in which the new data are incorporated (red with circular markers). The former two models correspond to those formally evaluated on 2000-2019 data in Section 4.

Given the extreme nature of the 2020 observations, to be able to run the basic DFM, which features both Gaussian disturbances and constant volatility, we need to adopt an

[37] An undesirable feature of our approach is that we are shutting down the uncertainty about the observation which we are imputing. One option could be to randomly draw from the forecast of the bridge regression, rather than taking the point estimate. This would be computationally expensive.

Figure 12: REAL-TIME ESTIMATES OF ACTIVITY AROUND THE PANDEMIC



Notes. Daily estimates of quarterly US real GDP growth with associated 68% HPD bands. The red line with circular markers comes from our Bayesian DFM estimated using the novel data series. The blue line with crossed markers represents the corresponding estimates from a model with identical specification but which uses only the traditional monthly and quarterly series. The green solid line is produced using a basic DFM.

ad-hoc procedure to censor outliers, used e.g. by Stock and Watson (2002b), 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. Moreover, the absence of heterogeneous lead-lag dynamics in the basic model hinders its ability to reflect the rebound in activity after the partial reopening. For these two reasons, the basic framework tracks economic developments with a full month of delay through the course of the pandemic. Instead, the presence of SV and fat tails in the full model means the whole data can be utilized, and the heterogeneous dynamics can pick up the rapid rebound in activity. Another important observation is how, as data for this period is released, the uncertainty around the estimates of the full model

increases quickly as the model revises upwards the volatility of the common factor.[38]

Adding the alternative data to the full model allows it to capture the decline in activity faster than when using only the traditional monthly data. The estimate of economic activity based on traditional data would only be slightly negative until April 15, when industrial production and retail sales data for the month of March was released, after which it registers a massive drop. In contrast, the information contained in the high-frequency proxies during March and early April, such as the sharp declines in credit card spending, mobility and restaurant reservations, provides the model with more timely signals about the sharp economic downturn in the wake of the first lockdown. While many observers in media and policy circles did pay close attention to these new data sources, our method of incorporating them into the Bayesian DFM provides a systematic and formal way to do so. By using a number of different series jointly, and extracting a signal about aggregate economic activity, the DFM brings together the potentially noisy information in each of them. After April 15, the estimates from the full model with and without the new data are relatively similar. We conclude that the alternative data were useful in capturing the decline in activity early, but the Bayesian DFM with traditional data tracked the rebound in a similar manner.

5.3 Discussion: modeling time series beyond 2020

The next years of modeling macroeconomic time series will be characterized by the challenge that the Great Lockdown remains part of the sample, in the same way that a binding Zero Lower Bound on interest rates became an important feature of the data after 2008. As we have discussed above, the recent crisis differs from usual recessions both in its time-series dynamics, with a dramatic fall followed by a rapid rebound, and in the cross-sectional dimension, with an alteration of the usual comovement between consumption, investment, and output. The magnitude of the shock means that the 2020

[38] This contrasts with the results obtained from models without SV, in which, fixing the forecast horizon, uncertainty always declines in response to new data, as discussed by Banbura and Modugno (2014).

observations will become dominant whenever using estimation techniques featuring constant volatility and Gaussian shocks, so the usual assumptions of factor models will lead to unreliable results. Excluding several quarters of data, implying that there is *nothing* to be learnt about macroeconomic dynamics from this episode, is unlikely to be an acceptable strategy, as emphasized by Primiceri and Tambalotti (2020) in the context of VARs, especially if macroeconomic volatility remains elevated as the recovery lingers. Even the traditional techniques for incorporating stochastic volatility will be problematic, as these are not meant to capture the discrete but transitory increases in variance observed only in some of the affected series. Therefore, understanding macroeconomic dynamics after 2020 will require a more explicit recognition that non-linearities and non-Gaussianities are a pervasive feature of modern business cycles. The framework developed in this paper is an approach that will likely remain beneficial even as new observations get appended to macroeconomic time series, and the COVID-19 pandemic moves into the rear-view mirror.

6 Conclusion

We propose a Bayesian DFM that incorporates low-frequency variation in the mean and variance of the variables, heterogeneous lead-lag responses to common shocks, and fat tails. The model is estimated via a fast and efficient algorithm that avoids non-linear computation. In a comprehensive evaluation exercise based on fully real-time, unrevised data, the nowcasting performance is substantially stronger than that of benchmark models and comparable or better than that of professional human forecasters. Our modeling innovations are material to understanding the evolution of activity during the COVID-19 recession and recovery, and will likely provide an advantageous framework for the study of macroeconomic time series that include the 2020 sample. Finally, we have laid out a method to incorporate newly available

high-frequency data into the DFM framework and shown that the new data provide important signals during March and April 2020.

References

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

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

ARUOBA, S. B. AND F. X. DIEBOLD (2010): "Real-time macroeconomic monitoring: Real activity, inflation, and interactions," *American Economic Review*, 100, 20–24.

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 (2013): "Now-Casting and the Real-Time Data Flow," in *Handbook of Economic Forecasting*, Elsevier, vol. 2, 195–237.

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.

BERAJA, M. AND C. K. WOLF (2021): "Demand Composition and the Strength of Recoveries," Unpublished manuscript, MIT.

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.

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.

BRUNNERMEIER, M. K., D. PALIA, K. A. SASTRY, AND C. A. SIMS (2020): "Feedbacks: financial markets and economic activity," *American Economic Review (Forthcoming)*.

CAMACHO, M. AND G. PEREZ-QUIROS (2010): "Introducing the euro-sting: Short-term indicator of euro area growth," *Journal of Applied Econometrics*, 25, 663–694.

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

CHAN, J. C. AND I. JELIAZKOV (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

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," NBER Working Papers 27431, NBER.

CIMADOMO, J., D. GIANNONE, M. LENZA, F. MONTI, AND A. SOKOL (2020): "Nowcasting with Large Bayesian Vector Autoregressions," Working Paper Series 2453, European Central Bank.

CLARK, T. AND M. MCCRACKEN (2013): "Advances in Forecast Evaluation," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 20, 1107–1201.

COGLEY, T. AND T. J. SARGENT (2005): "Drifts and volatilities: monetary policies and outcomes in the post WWII US," *Review of Economic dynamics*, 8, 262–302.

CÚRDIA, V., M. DEL NEGRO, AND D. L. GREENWALD (2014): "Rare shocks, great recessions," *Journal of Applied Econometrics*, 29, 1031–1052.

D'AGOSTINO, A., D. GIANNONE, M. LENZA, AND M. MODUGNO (2016): "Nowcasting Business Cycles: A Bayesian Approach to Dynamic Heterogeneous Factor Models," in *Dynamic Factor Models*, Emerald Publ., vol. 35 of *Advances in Econometrics*, 569–594.

DIEBOLD, F. X. (2020): "Real-Time Real Economic Activity Entering the Pandemic Recession," *Covid Economics*, 62, 1–19.

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.

FAUST, J. AND J. H. WRIGHT (2009): "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset," *Journal of Business & Economic Statistics*, 27, 468–479.

FERNÁNDEZ-VILLAVERDE, J., P. GUERRÓN-QUINTANA, K. KUESTER, AND J. RUBIO-RAMÍREZ (2015): "Fiscal volatility shocks and economic activity," *American Economic Review*, 105, 3352–84.

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.

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.

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.

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 (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

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," NBER Working Papers 27771, National Bureau of Economic Research.

LEWIS, D. J., K. MERTENS, J. H. STOCK, AND M. TRIVEDI (2020): "Measuring real activity using a weekly economic index," *Federal Reserve Bank of New York*, 920.

LUDVIGSON, S. C., S. MA, AND S. NG (2015): "Uncertainty and business cycles: exogenous impulse or endogenous response?" NBER Working Papers 21803.

MARCELLINO, M., M. PORQUEDDU, AND F. VENDITTI (2016): "Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility," *Journal of Business & Economic Statistics*, 34, 118–127.

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.

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

NAKAMURA, E. AND J. STEINSSON (2018): "High-frequency identification of monetary non-neutrality: the information effect," *The Quarterly Journal of Economics*, 133, 1283–1330.

PRIMICERI, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The 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*.

ROMER, C. D. AND D. H. ROMER (2000): "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90, 429–457.

ROSSI, B. AND T. SEKHPOSYAN (2019): "Alternative tests for correct specification of conditional predictive densities," *Journal of Econometrics*, 208, 638–657.

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. AND D. SONG (2020): "Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic," Working Papers 20-26, Philadelphia Fed.

SIMS, C. A. (2002): "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity*, 33, 1–62.

——— (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*, 351–409.

——— (2002a): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

——— (2002b): "Macroeconomic forecasting using diffusion indexes," *Journal of Business & Economic Statistics*, 20, 147–162.

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

——— (2017): "Twenty Years of Time Series Econometrics in Ten Pictures," *Journal of Economic Perspectives*, 31, 59–86.

Online Appendix to "Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data"

by Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

Contents

A Details on model and algorithm

A.1 Treatment of missing observations in vectorized Kalman Smoother

In this appendix we extend the results of Chan and Jeliazkov (2009) on the vectorized Kalman smoother for the case where there are missing observations. For simplicity we describe the algorithm for the case with fixed coefficients. Let use write the model as

$$\begin{aligned} \mathbf{y_t} &= \mathbf{H}\mathbf{x}_t + \eta_{\boldsymbol{t}}, & \eta_{\boldsymbol{t}} \sim N(0, \mathbf{R}) \\ \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{e}_t, & \mathbf{e}_t \sim N(0, \mathbf{Q}) \end{aligned}$$

Following Durbin and Koopman (2012), we can express the model in its vectorized form:

$$\begin{aligned} \mathbf{y} &= \overline{\mathbf{H}}\mathbf{x} + \eta, & \eta \sim N(0, \overline{\mathbf{R}}) \\ \overline{\mathbf{F}}\mathbf{x} &= \mathbf{x}^* + \mathbf{e}, & \mathbf{e} \sim N(0, \overline{\mathbf{Q}}) \end{aligned} \tag{14}$$

where $\mathbf{y}' = [\mathbf{y}_1', ..., \mathbf{y}_T']$, $\mathbf{x}' = [\mathbf{x}_1', ..., \mathbf{x}_T']$ and $\mathbf{x}^* = [\mathbf{x}_0', \mathbf{0}_{1\times k(T-1)}]$. $\mathbf{x}_0$ is the initialization for the state vector and its variance is initialized at $\mathbf{P}_0$. The system matrices of the measurement equations are $\overline{\mathbf{H}} = (\mathbf{I}_T \otimes \mathbf{H})$ and $\overline{\mathbf{R}} = (\mathbf{I}_T \otimes \mathbf{R})$ while those in the transition equation are:

$$\overline{\mathbf{F}} = \begin{bmatrix} \mathbf{I}_k & & & \\ -\mathbf{F} & \mathbf{I}_k & & \\ & \ddots & \ddots & \\ & & -\mathbf{F} & \mathbf{I}_k \end{bmatrix}, \quad \overline{\mathbf{Q}} = \begin{bmatrix} \mathbf{P}_0 & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_{T-1} \otimes \mathbf{Q}) \end{bmatrix}. \tag{15}$$

In order to deal with the the presence of a rank deficient system for the state variables, arising from the presence of multiple lags in the dynamics of the factors, we define $\mathbf{J}$ as a k-dimensional diagonal matrix with zeros corresponding to singular columns of the matrix $\mathbf{Q}$ and one for the nonsingular columns and let $\overline{\mathbf{J}}$=$(\mathbf{I}_T \otimes \mathbf{J})$. Therefore we define $\widetilde{\mathbf{F}} = \overline{\mathbf{J}}'\overline{\mathbf{F}}\overline{\mathbf{J}}$, $\widetilde{\mathbf{Q}} = \overline{\mathbf{J}}'\overline{\mathbf{Q}}\overline{\mathbf{J}}$ and $\widetilde{\mathbf{H}} = \overline{\mathbf{H}}\overline{\mathbf{J}}$. In order to deal with the presence of missing observations, we define $\boldsymbol{\Xi}$ as a diagonal matrix with ones corresponding to the existing elements in $\mathbf{y}$ and zero for the missing elements. Therefore we define as $\widetilde{\mathbf{R}}^{-1} = \boldsymbol{\Xi}'\overline{\mathbf{R}}^{-1}\boldsymbol{\Xi}$. Therefore one can sample from the state vector noting that $\mathbf{x} \sim \mathbf{N}(\varkappa, \mathbf{P})$

with

$$\begin{aligned}\mathbf{P} &= \mathbf{K}+\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\\ \varkappa &= \mathbf{P}\backslash\ \ \mathbf{K}\mathbf{F}\backslash\mathbf{x}^{*}+\mathbf{H}'\mathbf{R}^{-1}\mathbf{y}\end{aligned}$$

where $\mathbf{K}=\mathbf{F}'/\mathbf{Q}\mathbf{F}$.

A.2 Construction of the State Space System

For expositional clarity, we focus on the case with $\mathbf{B} = (1\ 0\ \ldots\ 0)'$ here. Recall that in our main specification we choose the order of the autoregressive dynamics in factor and idiosyncratic components to be $p = 2$ and $q = 2$, respectively. Let the $n \times 1$ vector $\tilde{\mathbf{y}}_t$, which contains n_q de-meaned quarterly and n_m de-meaned monthly variables (i.e. $n = n_q + n_m$), be defined as

$$\tilde{\mathbf{y}}_t = \begin{bmatrix} y^q_{1,t} \\ \vdots \\ y^q_{n_q,t} \\ y^m_{1,t} - \rho^m_{1,1} y^m_{1,t-1} - \rho^m_{1,2} y^m_{1,t-2} \\ \vdots \\ y^m_{n_m,t} - \rho^m_{n_m,1} y^m_{n_m,t-1} - \rho^m_{n_m,2} y^m_{n_m,t-2} \end{bmatrix},$$

so that the system is written out in terms of the *quasi-differences* of the monthly indicators. Given this re-defined vector of observables, we cast our model into the following state space form:

$$\begin{aligned}\tilde{\mathbf{y}}_\mathbf{t} &= \mathbf{H}\mathbf{x}_t + \tilde{\boldsymbol{\eta}}_\mathbf{t}, & \tilde{\boldsymbol{\eta}}_\mathbf{t} &\sim N(0, \tilde{\mathbf{R}}_\mathbf{t})\\ \mathbf{x}_t &= \mathbf{F}\mathbf{x}_{t-1} + \mathbf{e}_t, & \mathbf{e}_t &\sim N(0, \mathbf{Q}_t)\end{aligned}$$

where the state vector is defined as $\mathbf{x}'_t = [a_t, \ldots, a_{t-4}, f_t, \ldots, f_{t-4}, \mathbf{u}^{q\prime}_t, \ldots, \mathbf{u}^{q}_{t-4}{}']$. Setting $\lambda_1 = 1$ for identification, the matrices of parameters $\mathbf{H}$ and $\mathbf{F}$, are then constructed as shown below:

$$\mathbf{H} = \begin{bmatrix} & & \\ \mathbf{H}_a & \mathbf{H}_{\lambda_q} & \mathbf{H}_u \\ & \mathbf{H}_{\lambda_m} & \end{bmatrix},$$

where the respective blocks of $\mathbf{H}$ are defined as

$$\mathbf{H}_a = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 1 & \frac{2}{3} & \frac{1}{3} \\ \hline \mathbf{0}_{(n-1)\times 5} \end{bmatrix}, \qquad \mathbf{H}_{\lambda_q} = \begin{bmatrix} 1 & \lambda_2 & \dots & \lambda_{n_q} \end{bmatrix}' \times \begin{bmatrix} 1/3 & 2/3 & 1 & 2/3 & 1/3 \end{bmatrix},$$

$$\mathbf{H}_{\lambda_m} = \begin{bmatrix} \lambda_{n_q+1} - \lambda_{n_q+1}\rho^m_{1,1} - \lambda_{n_q+1}\rho^m_{1,2} & \mathbf{0}_{1\times 4} \\ \vdots & \vdots \\ \lambda_n - \lambda_n \rho^m_{n_m,1} - \lambda_n \rho^m_{n_m,2} & \mathbf{0}_{1\times 4} \end{bmatrix}$$

$$\mathbf{H}_u = \begin{bmatrix} \bar{\mathbf{H}}_u \\ \mathbf{0}_{n_m \times 5} \end{bmatrix}, \qquad \bar{\mathbf{H}}_u = \mathbf{1}_{n_q \times 1} \times \begin{bmatrix} 1/3 & 2/3 & 1 & 2/3 & 1/3 \end{bmatrix},$$

and

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} & & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 & & & \\ \vdots & & \mathbf{F}_{2+1} & & \vdots \\ \vdots & & & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{F}_{2+n_q} \end{bmatrix},$$

where the respective blocks of $\mathbf{F}$ are defined as

$$\mathbf{F}_1 = \begin{bmatrix} 1 & \mathbf{0}_{1\times 4} \\ \mathbf{I}_4 & \mathbf{0}_{4\times 1} \end{bmatrix} \qquad \mathbf{F}_2 = \begin{bmatrix} \phi_1 & \phi_2 & \mathbf{0}_{1\times 3} \\ \mathbf{I}_4 & & \mathbf{0}_{4\times 1} \end{bmatrix} \qquad \mathbf{F}_{2+j} = \begin{bmatrix} \rho^q_{j,1} & \rho^q_{j,2} & \mathbf{0}_{1\times 3} \\ \mathbf{I}_4 & & \mathbf{0}_{4\times 1} \end{bmatrix}$$

for $j = 1, ..., n_q$.

The error terms are denoted as

$$\begin{aligned} \tilde{\boldsymbol{\eta}}_t &= [\mathbf{0}_{1\times n_q}, \tilde{\boldsymbol{\eta}}^{\boldsymbol{m}\prime}_{\boldsymbol{t}}]' \\ \mathbf{e}_t &= \begin{bmatrix} v_{a_t} & \mathbf{0}_{4\times 1} & \epsilon_t & \mathbf{0}_{4\times 1} & \eta_{1,t} & \mathbf{0}_{4\times 1} & \dots & \eta_{n_q,t} & \mathbf{0}_{4\times 1} \end{bmatrix}', \end{aligned}$$

with covariance matrices

$$\tilde{\mathbf{R}}_\mathbf{t} = \begin{bmatrix} \mathbf{0}_{n_q\times n_q} & \mathbf{0}_{n_q\times n_m} \\ \mathbf{0}_{n_m\times n_q} & \mathbf{R}_t \end{bmatrix},$$

where $\mathbf{R}_t = diag(\sigma^2_{\eta^m_{1,t}}, \dots, \sigma^2_{\eta^m_{n_m,t}})$ and

$$\mathbf{Q}_t = diag(\omega^2_a, \mathbf{0}_{1\times 4}, \sigma^2_{\epsilon,t}, \mathbf{0}_{1\times 4}, \sigma^2_{\eta^q_{1,t}}, \mathbf{0}_{1\times 4}, \dots, \sigma^2_{\eta^q_{n_q,t}}, \mathbf{0}_{1\times 4}).$$

A.3 Details of the Gibbs sampler algorithm

Let $\boldsymbol{\theta} \equiv \{\boldsymbol{c}, \boldsymbol{\lambda}, \boldsymbol{\Phi}, \boldsymbol{\rho}, \omega_a, \omega_\varepsilon, \omega_{\eta_1}, \dots, \omega_{\eta_n}, \sigma_{o,1}, \dots, \sigma_{o,n}, v_1, \dots, v_n\}$ be a vector that collects the underlying parameters, where $\boldsymbol{\Phi}$ and $\boldsymbol{\rho}$ contain the parameters for factor and idiosyncratic components. The model is estimated using a Markov Chain Monte Carlo (MCMC) Gibbs sampling algorithm in which conditional draws of the latent variables, $\{a_t, f_t\}^T_{t=1}$, the parameters, $\boldsymbol{\theta}$, and the stochastic volatilities, $\{\sigma_{\varepsilon,t}, \sigma_{\eta_{i,t}}\}^T_{t=1}$ are obtained sequentially. The algorithm has a block structure composed of the following steps.

0. Initialization

The model parameters are initialized at arbitrary starting values $\boldsymbol{\theta}^0$, and so are the sequences for the stochastic volatilities, $\{\sigma^0_{\varepsilon,t}, \sigma^0_{\eta_{i,t}}\}_{t=1}^T$. The latent components $\mathbf{c}_t$, $\mathbf{o}_t$, and $\mathbf{f}_t$, are initialized by running the Kalman filter and smoother once conditional on the initialized parameters. Set $j = 1$.

1. Draw outlier component conditional on estimated common factors

Obtain a draw $\{o_{i,t}\}_{t=1}^T$ from $p(\{o_{i,t}\}_{t=1}^T|\boldsymbol{\theta}^{j-1}, a_t^{j-1}, f_t^{j-1}, \{\sigma^{j-1}_{\varepsilon,t}, \sigma^{j-1}_{\eta_{i,t}}\}_{t=1}^T, \mathbf{y})$ for each variable, $i = 1, \ldots, N$.

Conditioning on a_t^{j-1}, f_t^{j-1} and the loadings, λ_i^{j-1}, one can compute $\Delta y_{i,t} - c_{i,t} - \lambda_i(L)f_t$. Therefore, conditioning on $\boldsymbol{\rho}_i^{j-1}$, $\{\sigma^{j-1}_{\eta_{i,t}}\}_{t=1}^T$, and $\sigma^{j-1}_{o,i}$ and v_i^{j-1}, one can use the Kalman filter and simulation smoother to independently draw the outlier components. This step is independent for each of the variables in the system as such it can be parallelized.

2. Draw the parameters of the outlier component

For each variable, $i = 1, \ldots, N$, obtain a draw from of $\sigma_{o,i}$ from $p(\sigma_{o,i}|\{o^j_{i,t}\}_{t=1}^T, v_i^{j-1})$ and v_i from $p(v_i|\{o^j_{i,t}\}_{t=1}^T, \sigma^j_{o,i})$.

A fat-tailed distribution is easily obtained by a scale mixture, see Geweke (1993), Jacquier et al. (2002). Therefore, we can treat the scale mixture variable as a latent variable. Specifically, assume that $\psi_{i,t}$ is distributed i.i.d. inverse gamma, or that $v_i/\psi_{i,t} \sim \chi^2_{v_i}$, one has that $\sqrt{\psi_{i,t}} z_t \sim t_{v_i}(0, 1)$. Therefore, taking the sample $\{o^j_{i,t}/\sqrt{\psi^{j-1}_{i,t}}\}_{t=1}^T$ and posing an inverse-gamma prior $p(\sigma^2_{o,i}) \sim IG(s_{o,i}, \nu_{o,i})$ the conditional posterior of $\sigma^2_{o,i}$ is also drawn inverse-gamma distribution. We choose the scale $s_{o,i} = 0.1$ and degrees of freedom $\nu_{o,i} = 1$ for our the monthly variables. For the quarterly variables where only one in three data points is available, we choose a more conservative prior with $\nu_{o,i} = 30$ degrees of freedom.

Conditioning on $\sigma^j_{o,i}$, given the conjugate inverse gamma prior, the conditional posterior of $\psi_{i,t}|v$ is also an inverse gamma. A draw $\psi^j_{i,t}$ can therefore be obtained from $p(\psi^j_{i,t}|o^j_{i,t}, \sigma^j_{o,i}, v_i^{j-1}) \sim IG\left(\frac{v_i^{j-1}+1}{2}, \frac{2}{(o^j_{i,t}/\sigma^j_{o,i})^2+2}\right)$ The degree of freedom v_i are discrete with probability mass

proportional to the product of t distribution ordinates $p(v_i^j|o_{i,t}^j, \sigma_{o,i}^j) = p(v) \prod_{t=1}^{T} \frac{v^{-1/2}\Gamma(v+1/2)}{\Gamma(1/2)\Gamma(v/2)}(v + (o_{i,t}^j/\sigma_{o,i}^j)^2)^{-(v+1)/2}$, where $p(v)$ denotes the prior distribution for the degree of freedom. We use a weakly informative prior for v which we assume to follow a Gamma distribution $\Gamma(2, 10)$ discretized on the support [3; 40]. This prior was proposed and analyzed by Juárez and Steel (2010). The lower bound at 3 enforces the existence of a conditional variance for the outlier component.

3. Draw latent variables conditional on model parameters and SVs

Obtain a draw $\{a_t^j, f_t^j, \mathbf{u}_t^q\}_{t=1}^T$ from $p(\{a_t, f_t\}_{t=1}^T|\{o_{i,t}^j\}_{t=1}^T, \boldsymbol{\theta}^{j-1}, \{\sigma_{\varepsilon,t}^{j-1}, \sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T, \mathbf{y})$.

Having computed the outlier adjusted series as $\Delta y_{i,t}^{OA} = \Delta(y_{i,t} - o_{i,t})$, this step of the algorithm uses the state space representation described above, and produces a draw from the entire state vector $\mathbf{X}_t$ (which includes the long-run growth components, a_t, the common factor, f_t, and the idiosyncratic components of the quarterly variables, $\mathbf{u}_t^q$) using the precision filter as described in Section A.1. Like Bai and Wang (2015), we initialise the Kalman Filter step from a normal distribution whose moments are independent of the model parameters, in particular $\mathbf{X}_0 \sim N(0, 10^4\mathbf{I})$.

4. Draw the variance of the time-varying GDP growth component

Obtain a draw $\omega_a^{2,j}$ from $p(\omega_a^2|\{a_t^j\}_{t=1}^T)$.

Taking the sample $\{a_t^j\}_{t=1}^T$ drawn in the previous step as given, and posing an inverse-gamma prior $p(\omega_a^2) \sim IG(S_a, v_a)$ the conditional posterior of ω_a^2 is also drawn inverse-gamma distribution. We choose the scale $S_a = 10^{-3}$ and degrees of freedom $v_a = 1$.

5. Draw the autoregressive parameters of the factor VAR

Obtain a draw $\boldsymbol{\Phi}^j$ from $p(\boldsymbol{\Phi}|\{f_t^j, \sigma_{\varepsilon,t}^j\}_{t=1}^T)$.

Taking the sequences of the common factor $\{f_t^j\}_{t=1}^T$ and its stochastic volatility $\{\sigma_{\varepsilon,t}^{j-1}\}_{t=1}^T$ from previous steps as given, and posing a non-informative prior, the corresponding conditional posterior is drawn from the Normal distribution, see, e.g. Kim and Nelson (1999). In the more

general case of more than one factor, this step would be equivalent to drawing from the coefficients of a Bayesian VAR. Like Kim and Nelson (1999), or Cogley and Sargent (2005), we reject draws which imply autoregressive coefficients in the explosive region.

6. Draw the factor loadings and constant terms

Obtain a draw of $\boldsymbol{\lambda}^j$ and $\boldsymbol{c}^j$ from $p(\boldsymbol{\lambda}, \boldsymbol{c}|\rho^{j-1}, \{f_t^j, \sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T, \mathbf{y})$.

Conditional on the draw of the common factor $\{f_t^j\}_{t=1}^T$, the measurement equations reduce to n independent linear regressions with heteroskedastic and serially correlated residuals. By conditioning on $\boldsymbol{\rho}^{j-1}$ and $\sigma_{\eta_{i,t}}^{j-1}$, the loadings and constant terms can be estimated using GLS. When necessary, we apply linear restrictions on the loadings. In order to ensure the identification of the model, we set the loading of the GDP equation associated to the (contemporaneous) common factor to unity (Bai and Wang, 2015).

7. Draw the serial correlation coefficients of the idiosyncratic components

Obtain a draw of $\boldsymbol{\rho}^j$ from $p(\boldsymbol{\rho}|\lambda^{j-1}, \{f_t^j, \sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T, \mathbf{y})$.

Taking the sequence of the common factor $\{f_t^j\}_{t=1}^T$ and the loadings drawn in previous steps as given, the idiosyncratic components for the monthly variables can be obtained as $u_{i,t} = y_{i,t} - \lambda^j f_t^j$. For the quarterly variables, a draw of the idiosyncratic components has been obtained directly in Step 3. Given a sequence for the stochastic volatility of the i^{th} component, $\{\sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T$, the residual is standardized to obtain an autoregression with homoskedastic residuals whose conditional posterior can be drawn from the Normal distribution.

8. Draw the stochastic volatilities

Obtain a draw of $\{\sigma_{\varepsilon,t}^j\}_{t=1}^T$ and $\{\sigma_{\eta_{i,t}}^j\}_{t=1}^T$ from $p(\{\sigma_{\varepsilon,t}\}_{t=1}^T|\boldsymbol{\Phi}^j, \{f_t^j\}_{t=1}^T)$, and from $p(\{\sigma_{\eta_{i,t}}\}_{t=1}^T|\boldsymbol{\lambda}^j, \boldsymbol{\rho}^j, \{f_t^j\}_{t=1}^T, \mathbf{y})$ respectively.

Finally, we draw the stochastic volatilities of the innovations to the factor and the idiosyncratic components independently, using the algorithm of Kim et al. (1998), which uses a mixture of

normal random variables to approximate the elements of the log-variance. This is a more efficient alternative to the exact Metropolis-Hastings algorithm previously proposed by Jacquier et al. (2002). For the general case in which there is more than one factor, the volatilities of the factor VAR can be drawn jointly, see Primiceri (2005).

Increase j by 1 and iterate until convergence is achieved.

B Data series included in the analysis

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

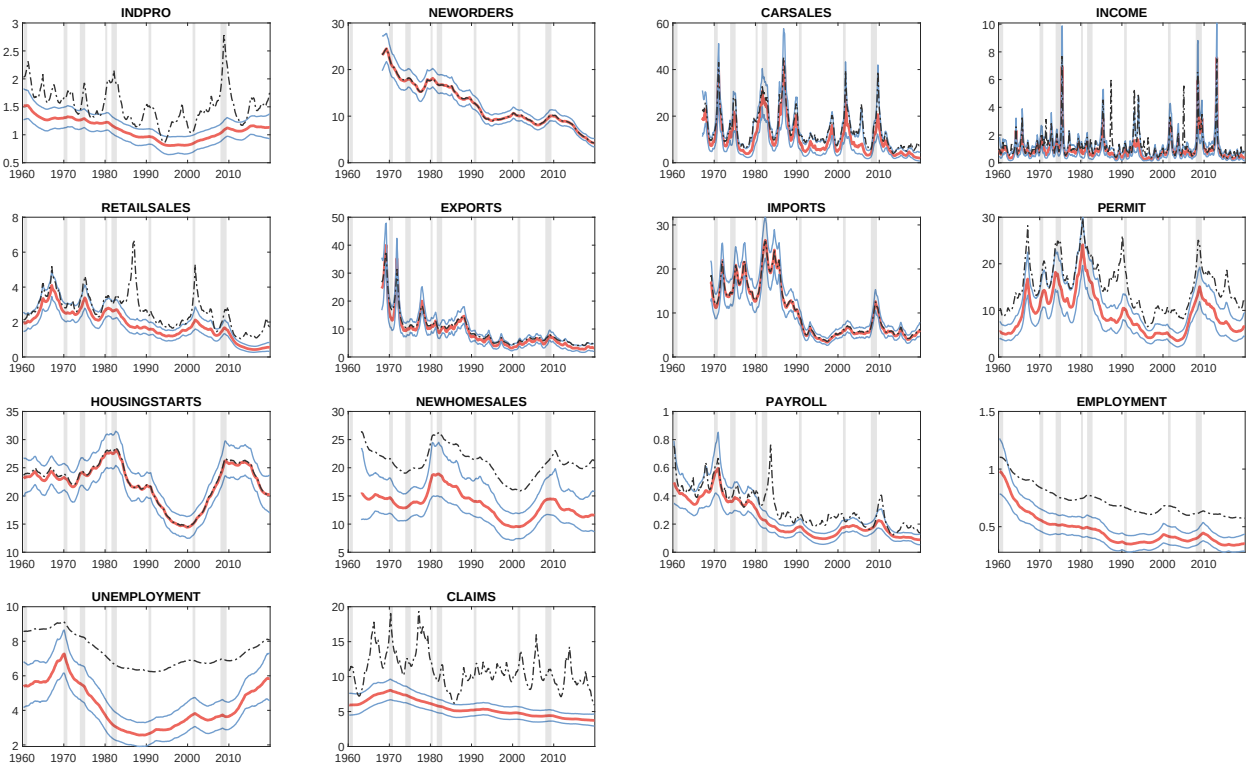
	Type	Start Date	Transform.	Lag
QUARTERLY TIME SERIES				
Real GDP	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real GDI	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Consumption (excl. durables)	Expenditure & Inc.	Q2:1947	% QoQ Ann	26
Real Investment (incl. durable cons.)	Expenditure & Inc.	Q2:1947	% QoQ Ann	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 in-sample results

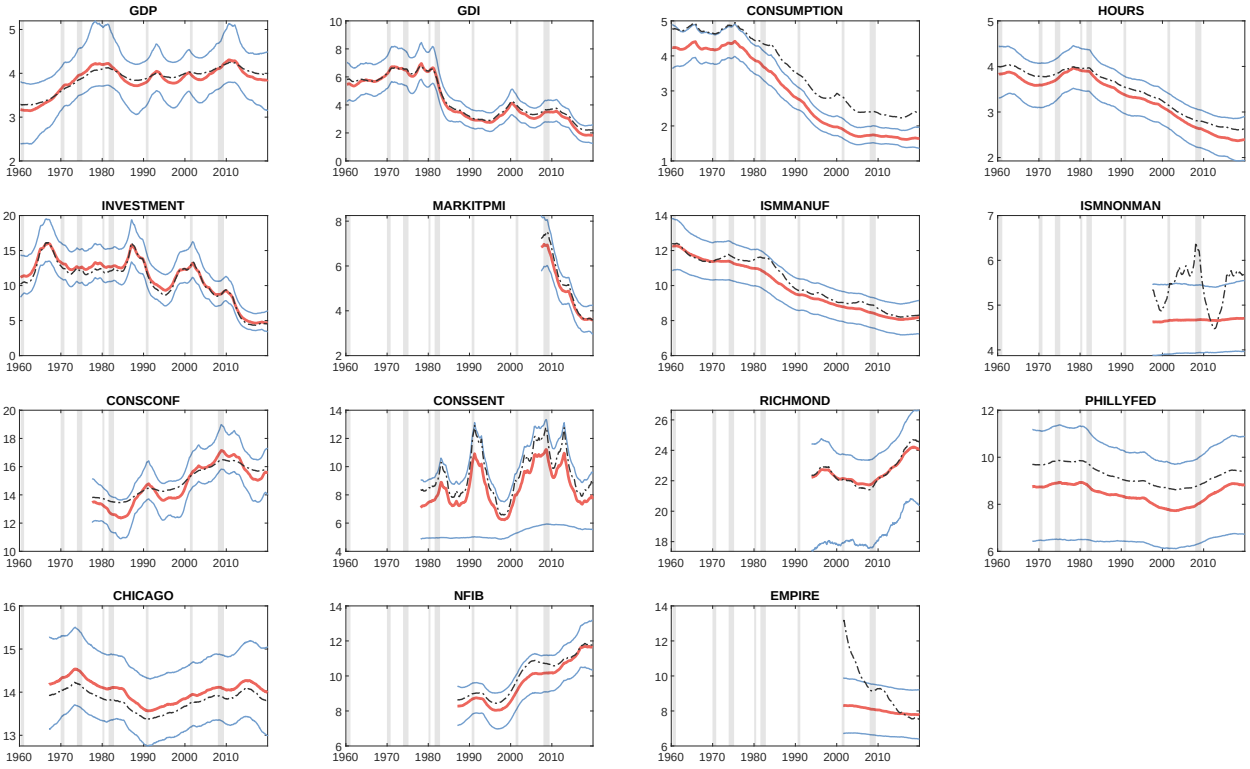
C.1 SV of idiosyncratic components

Figure C.1: POSTERIOR ESTIMATE OF SV OF MONTHLY HARD VARIABLES



Notes. Each panel presents the median (solid red), the 68% and the 90% (solid and dashed blue) posterior credible intervals of the volatility of the idiosyncratic component of different variables in our baseline DFM. For comparison, the black dashed-dotted line shows the median estimate for a version of the model that does not incorporate a Student-t component. Shaded areas represent NBER recessions. This figure contains the SV for the monthly hard variables in the data panel, while Figure C.2 presents the quarterly variables as well as the monthly soft variables. Table B.1 provides a full list of the individual data series with more details.

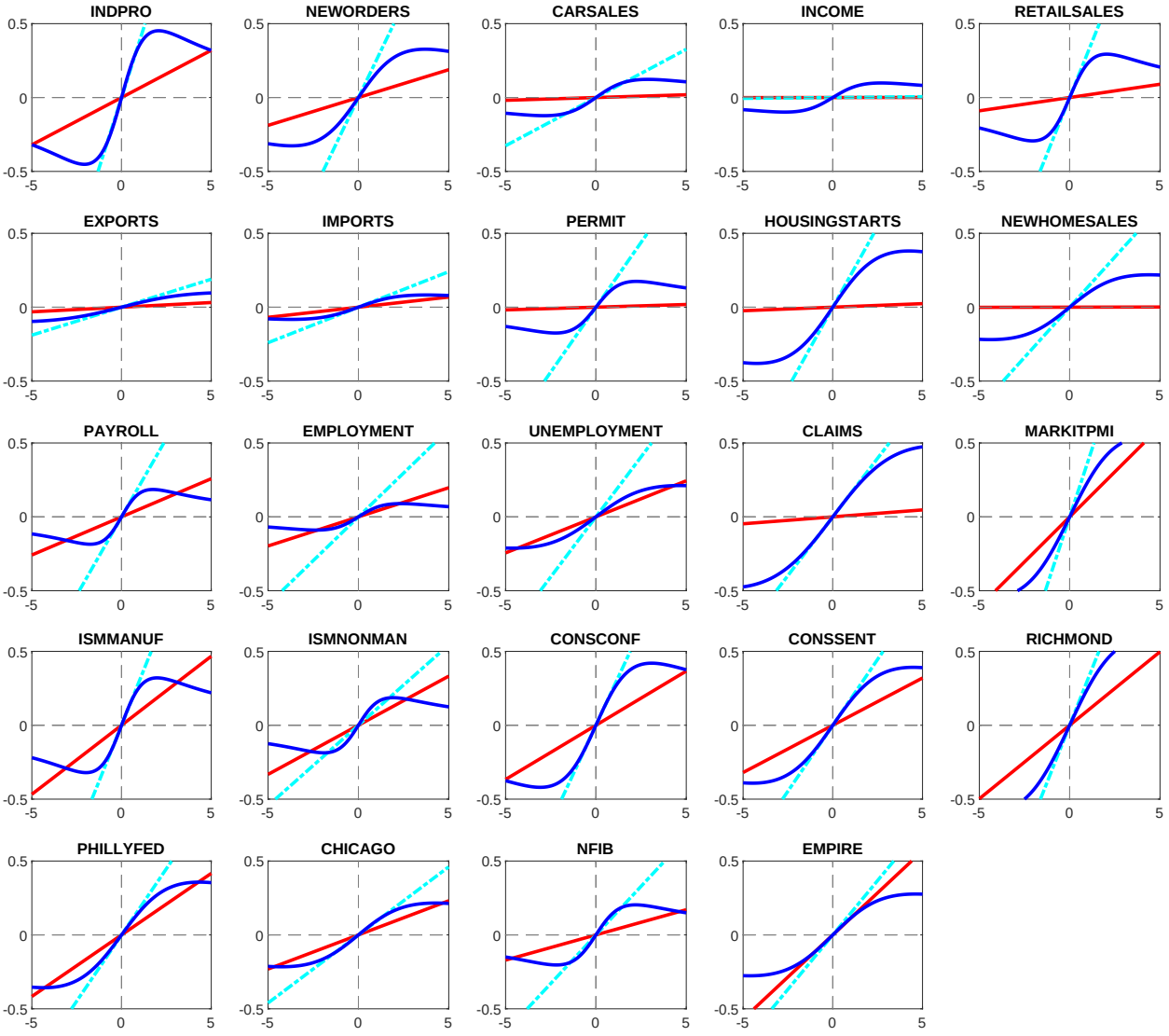
Figure C.2: POSTERIOR ESTIMATE OF SV OF QUARTERLY AND MONTHLY SOFT VARIABLES



Notes. Each panel presents the median (solid red), the 68% and the 90% (solid and dashed blue) posterior credible intervals of the volatility of the idiosyncratic component of different variables in our baseline DFM. For comparison, the black dashed-dotted line shows the median estimate for a version of the model that does not incorporate a Student-t component. Shaded areas represent NBER recessions. This figure contains the SV for the quarterly variables as well as the monthly soft variables in the data panel, while Figure C.1 presents the monthly hard variables. Table B.1 provides a full list of the individual data series with more details.

C.2 Influence function for all variables

Figure C.3: INFLUENCE FUNCTIONS FOR ALL VARIABLES



Notes. The panels of the figure plot the *influence functions* for all 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 red lines plot these influence functions in the Gaussian case with homogeneous dynamics. The dashed light blue lines represent the Gaussian case when we allow for heterogeneous dynamics, that is, lags of the factor in the measurement equation. The dark blue lines correspond to our full model, where we also allow for the Student-t components. As shown in equations (10) - (12) and explained in the main text, the model with fat tails allows these functions to be nonlinear and nonmonotonic.

D Details on setup for the real-time out-of-sample evaluation

D.1 Construction of the real-time database

Macroeconomic data are 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, we collect *unrevised* 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:

1. 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.

2. Some 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.

3. 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 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.

D.2 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: it is sufficient take 7,000 iterations of the Gibbs sampler presented in Section 2.4 and Appendix A.3, discarding the first 2,000 as burn-in draws.

One such run takes approximately 30 minutes using a high-performance desktop computer, so the entire exercise across all vintages and different versions of the model would take several months.[1] We leverage the possibilities of massively parallelized cloud computation. We have integrated our codes with the Amazon Elastic Compute Cloud (Amazon EC2), which allows us to compute up to 2,500 runs of the algorithm simultaneously. This drastically reduces the amount of time required to asses the real-time performance of our Bayesian DFM and several benchmark models over the the period of twenty years.

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

E Additional forecast evaluation results

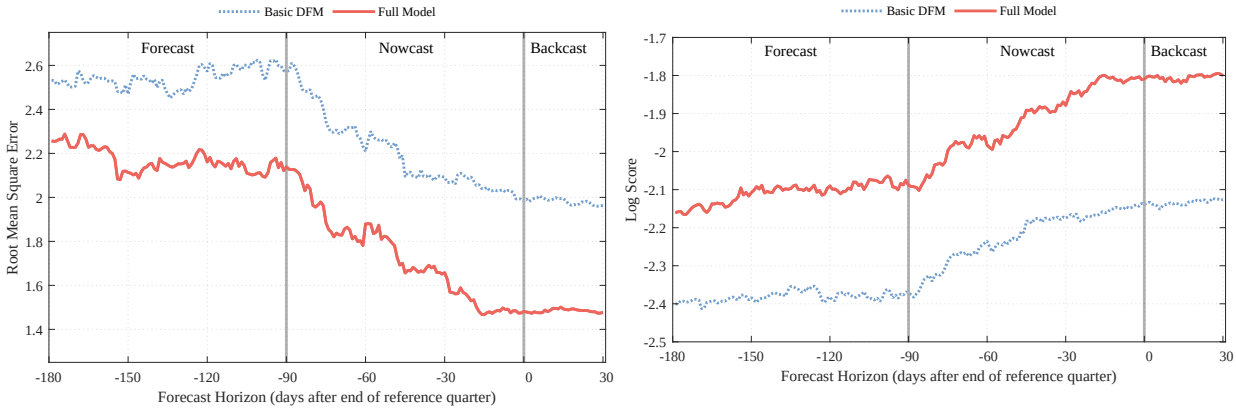
E.1 Forecast evaluation as the data flow arrives

Figure E.1 shows the accuracy of our Bayesian DFM relative to a basic DFM in predicting the third release of GDP. As in the main text, the basic 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. (2013). For these two models, the figure essentially opens up the results in Table 1 in the main text by providing a more continuous evolution of the point and density forecast evaluation metrics across the horizon. Specifically, we evaluate this accuracy starting 180 days before the end of the reference quarter (a forecast), as the reference quarter unfolds (90 to 0 days, 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 Section E.3. An analogous figure that breaks down the performance into the individual model components and compares it to the NY Fed model can be found in Sections E.4 and E.5.

Panel (a) shows that RMSE of both models declines as forecast horizon gets shorter and information contained in monthly indicators becomes available. The RMSE of the 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. (2013), an important property of an efficient nowcast is that the forecast error declines monotonically as new information arrives. 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.

Figure E.1: GDP FORECAST EVALUATION AS THE DATA ARRIVES

(a) Point forecast accuracy: RMSE **(b)** Density forecast accuracy: Log score



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 forecast 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.

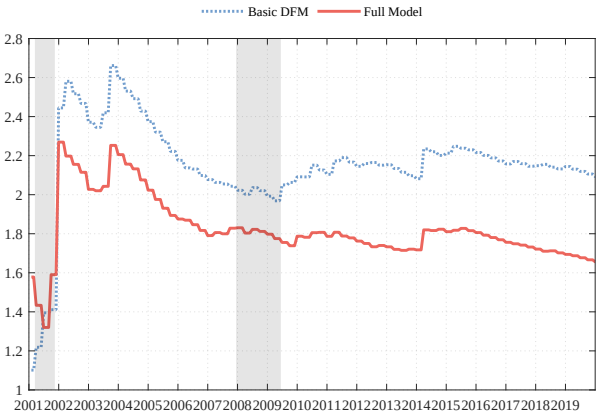
Panel (b) turns to density forecasts, so evaluates the accuracy of the entire predictive distribution, instead of focusing exclusively on its center. They are used to predict unusual developments or tail risks, such as the probability of a recession or a strong recovery given current information. Our 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 (we use an alternative metric further below). 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.

E.2 Forecast evaluation through time

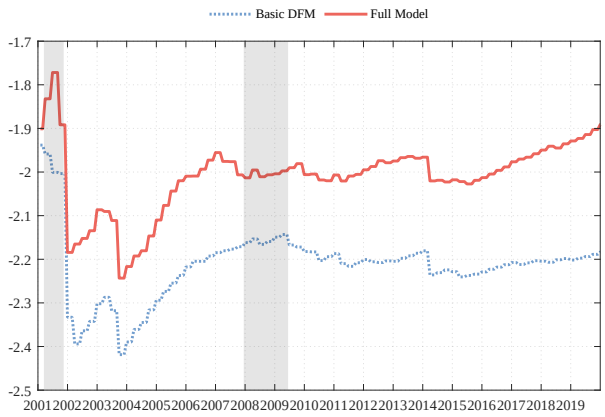
The results in Figure E.1 document the average performance over the period 2000-2019. It is useful to examine whether the out-performance of the full model is stable over time or due to a few special periods. In Figure E.2 we present, for a fixed forecast horizon, the relevant loss function computed 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 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 examine in the main text.

Figure E.2: GDP FORECAST EVALUATION THROUGH TIME

(a) Point forecast accuracy: RMSE



(b) Density forecast accuracy: Log score



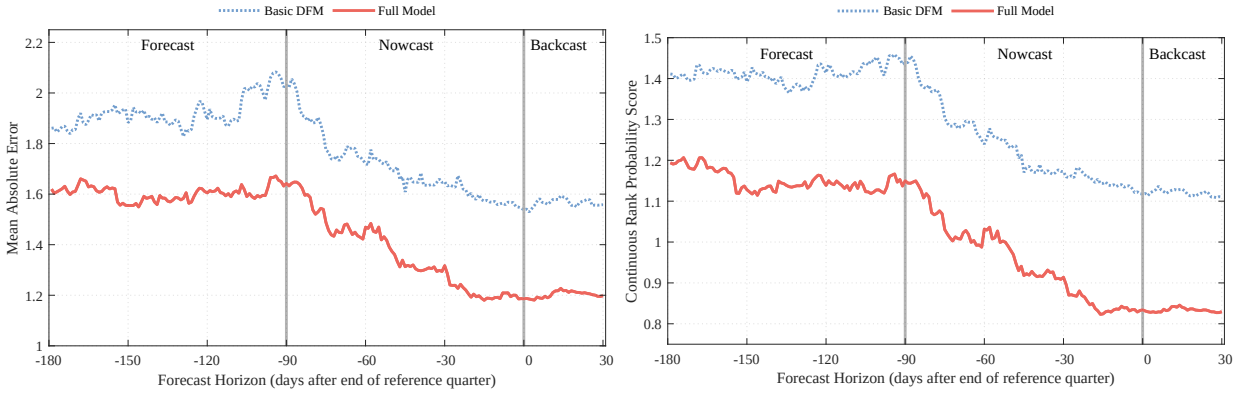
Notes. In both panels, the horizontal axis indicates the time of the evaluation sample 2000-2019. Panel (a) plots the rolling RMSE of the full model (solid red line) as well as the basic DFM (dotted blue) through time. A more accurate forecast implies a lower RMSE. Panel (b) displays the analogous rolling 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.

E.3 Alternative metrics for forecast evaluation

This Appendix examines the robustness of our formal forecast evaluation for alternative evaluation metrics. Figures E.3 and E.4 present the results shown in Figures E.1 and E.2, using alternative evaluation metrics. Figure E.3 focuses on the evaluation across horizons (as the data flow arrives) and Figure E.4 on the evaluation through time. In both cases, panel (a) presents point forecast evaluation results for the mean absolute error (MAE) rather the the RMSE. Panel (b) turns to density forecast evaluation and applies the continuous rank probability score (CRPS) instead of the Log score. It is evident that the the conclusions drawn from our evaluation exercise are broadly robust to using different evaluation metrics.

Figure E.3: FORECAST EVALUATION AS THE DATA FLOW ARRIVES (ALTERNATIVE METRICS)

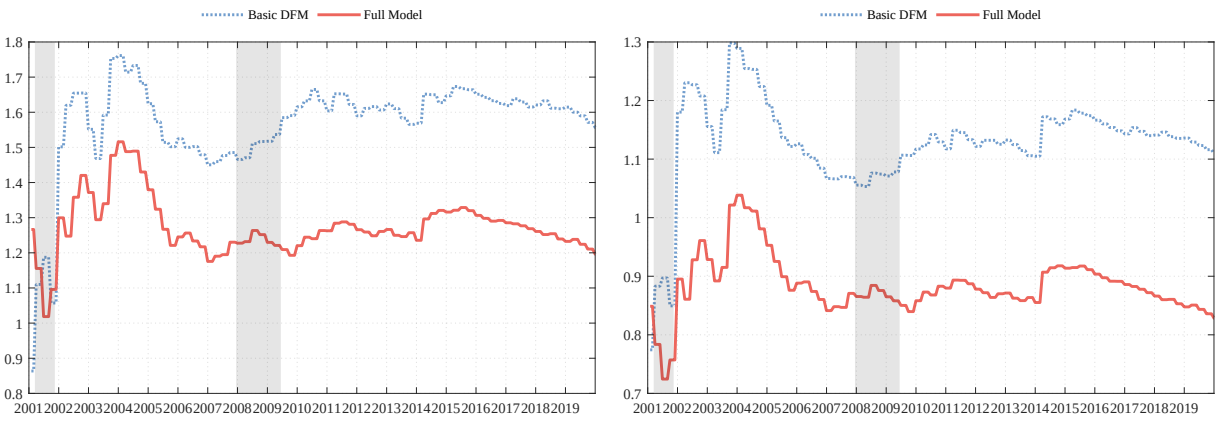
(a) Point forecast accuracy: MAE **(b)** Density forecast accuracy: CRPS



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 forecast 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 E.4: FORECAST EVALUATION THROUGH TIME (ALTERNATIVE METRICS)

(a) Point forecast accuracy: MAE **(b)** Density forecast accuracy: CRPS



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 time. 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.

E.4 Relative contribution of different components

Table E.1 reports the evaluation exercise for models of increasing complexity: first, a simple autoregressive (AR) model of order 1 on quarterly data; second, the basic DFM; 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. In brackets we report the p-value of a formal test against the AR(1) model.

The table shows that 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.[2] Another substantial improvement comes from the addition of heterogeneous dynamics. 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.

For both point and density forecasting, the fat tails appear to lead only to a small improvement in the forecast accuracy relative to the model that has the heterogeneous dynamics. Not that that this evaluation is carried out over the 2000 to 2019 period for which GDP did not feature substantial outliers As we show in Section 5 of the main text, our modeling of fat tails plays a critical role during the COVID-19 crisis.

[2] As we point out in more detail further below, 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.

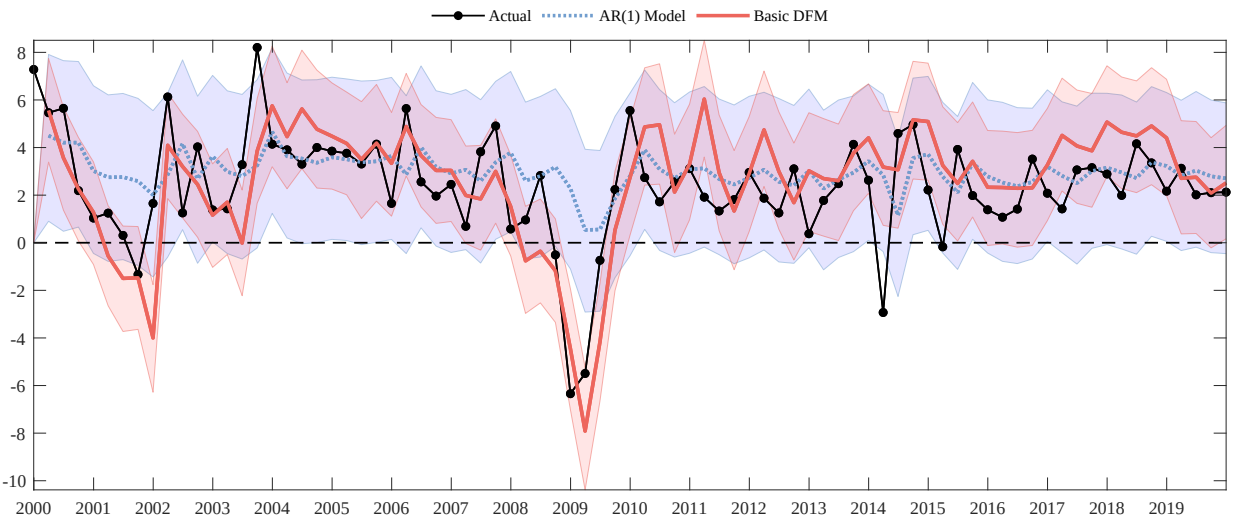
Table E.1: FORECASTING PERFORMANCE OF INTERMEDIATE MODEL VERSIONS

Point forecasting: RMSE	**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]
Density Forecasting: Log Score	**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 GDP data; the basic DFM; a DFM 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 components ("fat tails"). We report the performance in RMSE (top panel) and Log score (bottom panel) across various forecasting horizons. The p-value of a formal test against the AR(1) model reported in brackets. The test is computed using the Diebold and Mariano (1995)'s statistic with small-sample correction as suggested by Clark and McCracken (2013).

To highlight one important aspect of the discussion around Table 1 in the main text and Table E.1 above, Figure E.5 shows the nowcasts obtained from the AR(1) model and the basic DFM over time, together with the actual (advance) release of GDP. The forecast horizon is fixed to the day before the release. The figure shows AR(1) nowcast is generally relatively well centered around the release – which leads to reasonably low RMSEs – with the important exception of the two recessions in the sample, 2001 and 2007/08. While our Bayesian DFM is more accurate than the AR(1) in any case – so the key conclusions from our evaluations exercise are not affected by this pattern – this observation highlights that a relative comparison between the AR(1) and the basic DFM (as provided by Table E.1) may be somewhat unfair. Arguably, a model to track economic activity which performs extremely poorly in recessions should not be selected.

Figure E.5: MODEL NOWCASTS AND PUBLISHED GDP GROWTH COMPARED



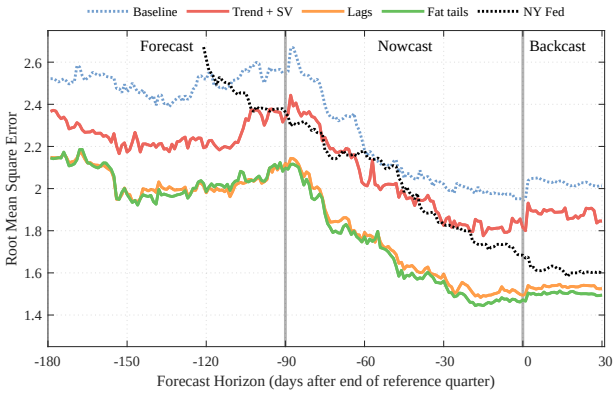
Notes. The black line represents the time series of actual real GDP growth in the United States. The blue line plots the nowcasts for real GDP growth based on information up to this point our basic DFM (red) and the AR(1) model (blue). The blue and red shaded areas indicate the corresponding density around the nowcasts.

E.5 More detailed comparison with NY FED Staff Nowcast

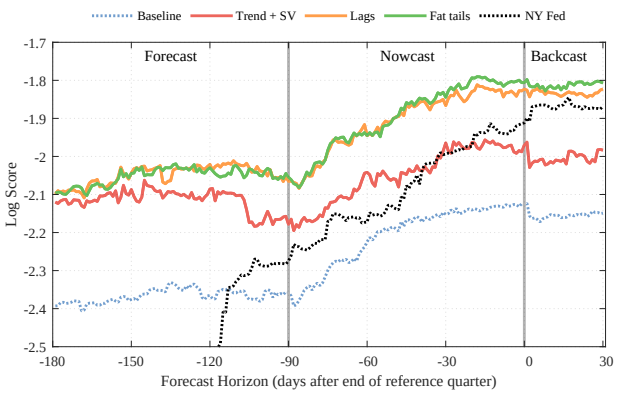
This appendix provides a more detailed comparison of the forecasting performance of our Bayesian DFM, as well as its individual novel components, relative to the New York Fed Staff Nowcast. Figure E.6 provides the evolution of RMSE and Log score over the forecasts horizon (as the data arrives). Panels (a) and (b) essentially correspond to the information in the two panels of Table 1 in the main text, but provide a graphical representation over a continuous evolution of the forecast horizon. Furthermore, relative to Table 1 the figure breaks down the performance of our model into the contribution of the individual components: trends and SV, heterogeneous dynamics and fat tails (similar to what we do in Table E.1 above). The figure provides a rich picture of forecasting performance of the different models across horizons. Overall, it tells the same story as our evaluation in the main text.

Figure E.6: MODEL COMPARISON AS THE DATA ARRIVES

(a) Point forecast accuracy: RMSE



(b) Density forecast accuracy: Log score

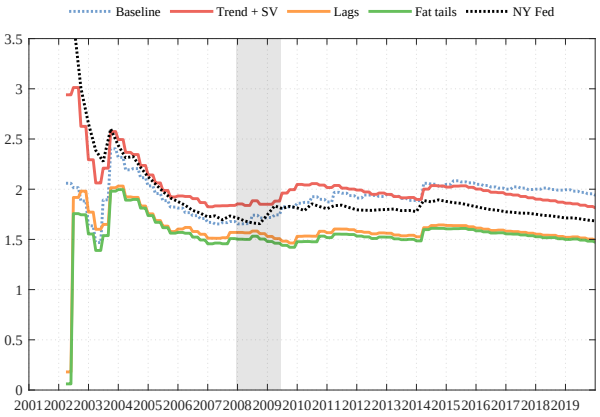


Notes. In both panels, the horizontal axis indicates the forecast horizon, expressed as the number of days to the end of the reference quarter of a given GDP release. Thus, from the point of view of the GDP forecaster, forecasts produced 180 to 90 days before the end of a given quarter are a forecast of the next quarter; forecasts at a 90-0 day horizon are nowcasts of the current quarter, and the forecasts produced 0-30 days after the end of the quarter are backcasts of the last quarter. Panel (a) plots the RMSE of over this horizon for the following models: the basic DFM; a version with time-varying long-run growth and SV (as in Antolin-Diaz et al., 2017) (labeled "trend & SV"); a version which adds, on top, the heterogeneous dynamics ("Lead/Lag"); the full model with the t-distributed component ("fat tails"); the NY Fed's nowcasting model. A more accurate forecast implies a lower RMSE. Panel (b) displays the analogous evolution of the log score of the different models, a measure of density forecast accuracy, which takes a higher value when a forecast is more accurate. Note that the New York Fed's model is a frequentist model that does not produce density forecasts. We construct the associated density forecasts by resampling from past forecast errors as in Bok et al. (2018).

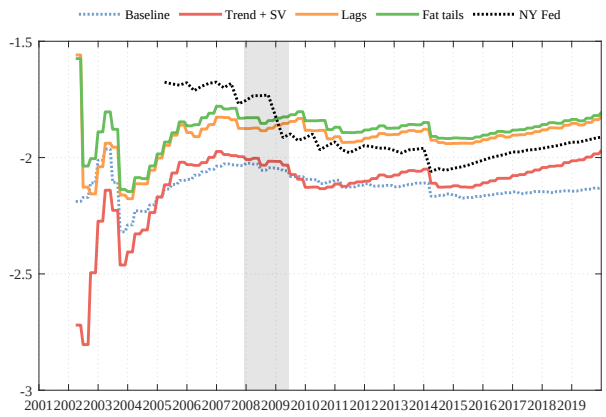
The results shown in Figure E.6 document the average performance over the period 2000-2019. Figure E.7 instead presents, for a fixed forecast horizon, the relevant loss function computed recursively *over time*. In other words, this figures extends the analysis in Appendix E.2 to a finer breakdown of model versions and includes the NY Fed's nowcasting model.

Figure E.7: MODEL COMPARISON THROUGH TIME

(a) Point forecast accuracy: RMSE



(b) Density forecast accuracy: Log score



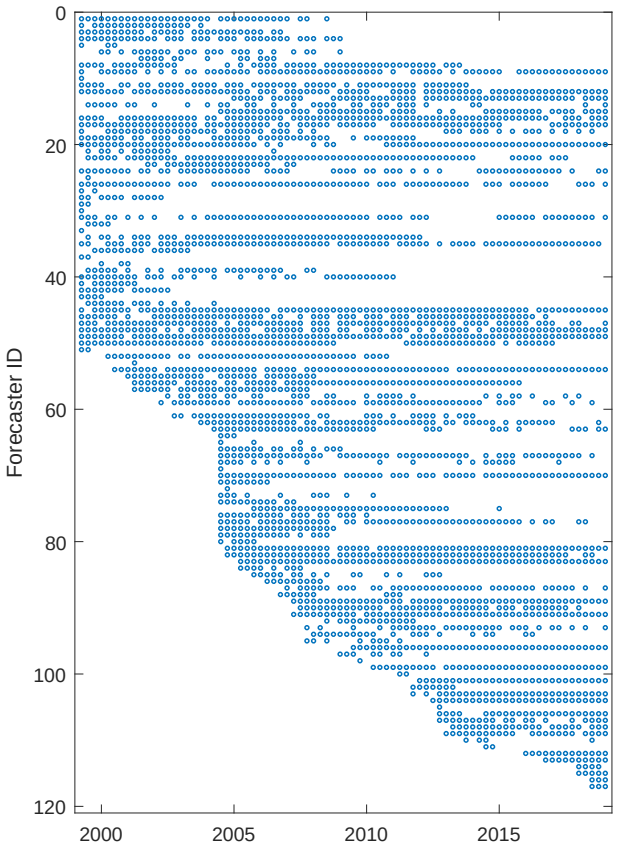
Notes. In both panels, the horizontal axis indicates the time of the evaluation sample 2000-2019. Panel (a) plots the rolling RMSE through time for the following models: the basic DFM; a version with time-varying long-run growth and SV (as in Antolin-Diaz et al., 2017) (labeled "trend & SV"); a version which adds, on top, the heterogeneous dynamics ("Lead/Lag"); the full model with the t-distributed component ("fat tails"); the NY Fed's nowcasting model. A more accurate forecast implies a lower RMSE. Panel (b) displays the analogous rolling evolution of the log score of the different 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. Note that the New York Fed's model is a frequentist model that does not produce density forecasts. We construct the associated density forecasts by resampling from past forecast errors as in Bok et al. (2018).

E.6 Details on using the Survey of Professional Forecasters

In Section 4.3 we have presented a comparison of our model to individual forecasts from the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia (see Croushore and Stark, 2019, for additional details). In this appendix we provide some additional background on the construction of the data required for that comparison, as well some additional analysis of the relative performance of our model for multiple period forecasts.

Since the Survey of Professional Forecasters is released at or before the 15th of the middle month in each quarter, we evaluate our model at horizon -45. In order to compare our model with individual participants in the SPF, we have to deal with the fact that individual responses are not continuously present in the SPF. This arises because the membership participation in the SPF is continuously updated by the Federal Reserve Bank of Philadelphia, and over time new forecasters are included in the sample while old one are excluded. Figure E.8 highlights the availability of individual respondent to SPF. Therefore to provide a reliable comparison in the main text we compare each individual forecasts on matched sample (i.e. comparing our model and the individual forecasters evaluating the RMSE only for the quarters when the individual forecasts are available). With macroeconomic volatility changing over the sample this guarantees that the comparison of the forecasts is on a like-for-like basis. The other challenge we face is which of the forecasts to include, in particular a balance needs to be stuck between the keeping the sample as large as possible yet making sure that the individual respondents have been present in the survey for a period long enough so that one can reliably evaluate their average forecast performance avoiding that this is unreasonably affected by single instances of good or bad luck. Taking both considerations into account led us to include only forecasters in the evaluation that are included in at least half of the sample (therefore the RMSE is computed on at least 40 forecasts). With this rule of thumb we end up with 40 individual forecasters, where their participation is reasonably equally spread over the out of sample evaluation window.

Figure E.8: AVAILABILITY OF INDIVIDUAL SPF FORECASTS

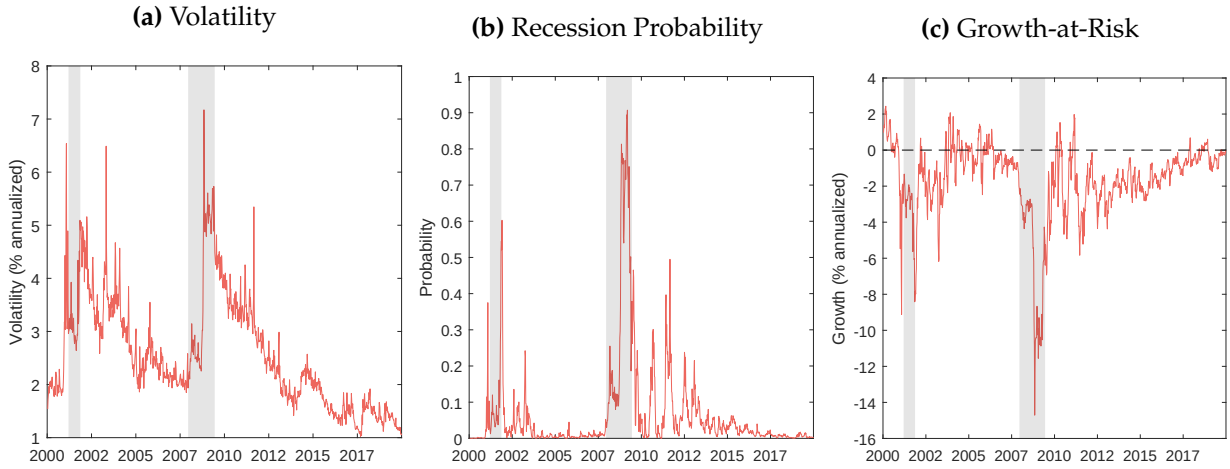


Notes. One line in the plot represents one individual forecaster ID. A blue dot indicates that the given forecaster has participated in the survey at a given point in time, while a white blank indicates the forecaster has not participated.

E.7 Real-time assessment of activity, uncertainty, tail risks

Our model can be used to derive 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.[3]

Figure E.9: 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 estimate. 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.

The probabilistic nature of our model allows us to compute additional statistics related to uncertainty and risk around our daily estimate of economic activity. In particular, Figure E.9 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 our daily estimate of activity. 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

[3]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 estimated values for underlying quarterly GDP, where the weights are the proportional to the number of days separating τ from the end of the quarter.

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 gives us confidence that they can be relied upon in practice.

References

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

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 (2013): "Now-Casting and the Real-Time Data Flow," in *Handbook of Economic Forecasting*, Elsevier, vol. 2, 195–237.

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.

CHAN, J. C. AND I. JELIAZKOV (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

CLARK, T. AND M. MCCRACKEN (2013): "Advances in Forecast Evaluation," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 20, 1107–1201.

COGLEY, T. AND T. J. SARGENT (2005): "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S," *Review of Economic Dynamics*, 8, 262–302.

CROUSHORE, D. AND T. STARK (2019): "Fifty Years of the Survey of Professional Forecasters," *Economic Insights*, 4, 1–11.

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

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

GEWEKE, J. (1993): "Bayesian Treatment of the Independent Student-t Linear Model," *Journal of Applied Econometrics*, 8, S19–S40.

JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (2002): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 20, 69–87.

JUÁREZ, M. A. AND M. F. STEEL (2010): "Model-based clustering of non-Gaussian panel data based on skew-t distributions," *Journal of Business & Economic Statistics*, 28, 52–66.

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.

PRIMICERI, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *The Review of Economic Studies*, 72, 821–852.