DISCUSSION PAPER SERIES

DP17800

ADVANCES IN NOWCASTING ECONOMIC ACTIVITY: THE ROLE OF HETEROGENEOUS DYNAMICS AND FAT TAILS

Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

INTERNATIONAL MACROECONOMICS AND FINANCE AND MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

ADVANCES IN NOWCASTING ECONOMIC ACTIVITY: THE ROLE OF HETEROGENEOUS DYNAMICS AND FAT TAILS

Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

Discussion Paper DP17800
Published 11 January 2023
Submitted 28 December 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

ADVANCES IN NOWCASTING ECONOMIC ACTIVITY: THE ROLE OF HETEROGENEOUS DYNAMICS AND FAT TAILS

Abstract

A key question for households, firms, and policy makers is: how is the economy doing now? This paper develops a Bayesian dynamic factor model that allows for nonlinearities, heterogeneous lead-lag patterns and fat tails in macroeconomic data. Explicitly modeling these features changes the way that different indicators contribute to the real-time assessment of the state of the economy, and substantially improves the out-of-sample performance of this class of models. In a formal evaluation, our nowcasting framework beats benchmark econometric models and professional forecasters at predicting US GDP growth in real time.

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

Keywords: Nowcasting, Dynamic factor models, Real-time data

Juan Antolin-Diaz - jantolindiaz@london.edu
London Business School

Thomas Drechsel - drechsel@umd.edu
University of Maryland College Park and CEPR

Ivan Petrella - ivan.petrella@wbs.ac.uk
Warwick Business School, University of Warwick and CEPR

Acknowledgements
This paper supersedes an earlier working paper circulated under the title ``Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data''. We thank conference and seminar participants at the NBER Summer Institute, 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 Society for Economic Dynamics Annual Meeting, 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, the European Central Bank, the Federal Forecasters Consortium, 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, Domenico Giannone and Mark Watson 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.

1 Introduction

Assessing macroeconomic conditions in real time is challenging. 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. Moreover, most macroeconomic series are revised over time, and many have become available only recently. Dynamic factor models (DFMs) are the workhorse tool to address these challenges. They exploit the idea that the variation in a large number of macroeconomic time series can be summarized by a few common factors (Sargent and Sims, 1977; Stock and Watson, 1989, 2002b; Forni et al., 2003). Based on the comovement among the time series, one obtains a real-time assessment of economic conditions and can construct *nowcasts* of GDP. Economic indicators that are released in a timely manner are particularly valuable for this purpose. Others are less useful since the information content of their release is captured by the data that is already available.

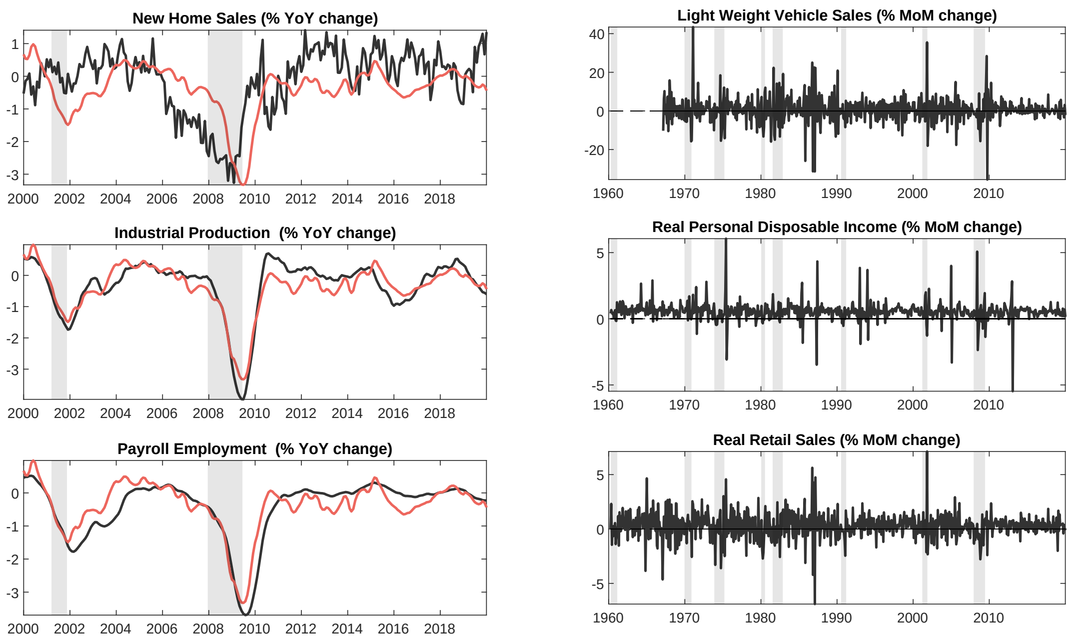
This paper shows that explicitly modeling two features of macroeconomic data typically ignored in standard DFM specifications –heterogeneous dynamics in the loadings to common factors, and fat tailed innovations in the idiosyncratic components– changes the way that different indicators contribute to the real-time estimation of the factors, and substantially improves the out-of-sample nowcasting performance of this class of models. Figure 1 illustrates these two features of the data. Panel (a) compares the annual growth rate of selected indicators to that of GDP. Lead-lag dynamics are pervasive in the data, with investment and durable spending leading GDP and labor variables lagging it. Panel (b) reveals 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, the above features of the data are often left unmodeled, with benchmark DFM specifications in the literature (Banbura et al., 2013; Antolin-Diaz et al., 2017; Bok et al., 2018) assuming homogeneous responses to common shocks and conditionally Gaussian shocks.

The methodological contribution of this paper is a fast and efficient algorithm to approximate the posterior distribution for DFM models featuring dynamic heterogeneity and additive fat-tailed components. Our algorithm can handle non-linearities and non-Gaussianities without

Figure 1: SALIENT FEATURES OF THE MACROECONOMIC DATA FLOW

(a) Heterogeneous lead-lag dynamics

(b) Fat tails



Notes. Panel (a) 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 (b) 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.

losing the intuition and computational ease of Kalman Filtering and Gibbs sampling. Building on ideas from Moench, Ng, and Potter (2013), we propose a hierarchical structure that avoids large state-spaces. 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.

The explicit modeling of heterogeneous dynamics and fat tails changes the way that information contained in macroeconomic indicators is interpreted and utilized in the nowcasting process by the DFM. In particular, it dramatically changes the weight given by the model to "hard" indicators –such as car sales, orders of durable goods, or housing starts– relative to "soft" indicators such as business surveys. The former are considered key indicators and play a central role in assessments of economic activity by policymakers and financial market participants, but are almost discarded by the standard model in favor of the latter. The literature has attributed this to the more timely nature of surveys, which are published weeks in advance of the hard data (see e.g. Banbura et al., 2013). Instead, we show that hard and soft data exhibit very different patterns of responses to common shocks, and that hard data are more prone to fat-tailed outliers. This introduces misspecification in the standard model which our model addresses, leading to a more balanced contribution of different series to the real-time estimation of the factors, even in light of the differences in publication delay.[1]

Our Bayesian DFM with heterogeneous dynamics and fat tails outperforms alternative models at real-time predictions of GDP growth, and improves upon survey expectations of professional forecasters. In our main empirical application, we estimate the model using data going back to 1947, construct daily predictions of US GDP growth over the period 2000-2019, and formally evaluate their accuracy using real-time unrevised vintage data. Our model delivers a large and highly significant improvement relative to various model benchmarks: a basic DFM, a DFM extended to include time-varying long-run growth and stochastic volatility (Antolin-Diaz, Drechsel, and Petrella, 2017) and the nowcasting model maintained by the New York Fed. Both point and density forecasting improve statistically and economically. Our model's nowcasts are also 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 Greenbook projections reveals that our model forecasts are as accurate early in the quarter, with the Greenbook taking an advantage later on.[2] 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, can be competitive with each other, though at very short horizons the Fed appears to display an informational advantage (Romer and Romer, 2000; Nakamura and Steinsson, 2018).[3]

In an additional empirical application, we analyze the recession of spring 2020 and the subsequent recovery, which pose unique challenges to macroeconometric models (see, e.g.,

[1]We formally examine the relation between modeling lead-lag dynamics in response to a common factor and alternative modeling choices based on multiple factors, and highlight the advantages of our approach.

[2]Human forecasters are a high benchmark for econometric models, 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).

[3]At longer horizons, both SPF surveys and FOMC projections display a noticeable upward bias in the last decade. We show that the time-varying long-run growth component of our DFM, proposed in Antolin-Diaz et al. (2017), eliminates this upward bias.

Lenza and Primiceri, 2020; Ng, 2021). By allowing for lead-lag patterns and fat-tails, our framework copes with the enormous swings in the data during this period without the need for any manual outlier adjustments or trimming of the data. The model tracks in real time the unprecedented fall in economic activity, and the quick but partial rebound. We discuss why its performance in producing accurate quarterly GDP nowcasts for 2020:Q2 and 2020:Q3 nevertheless remains limited.

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". In Antolin-Diaz, Drechsel, and Petrella (2017) we made the case for allowing both shifts in long-run growth and changes in volatility.[5] Here we develop the DFM framework further to include heterogeneous dynamics and student-t distributed outliers, and formally study their importance for the nowcasting process. We are the first to investigate the introduction of student-t distributed outliers within DFMs. In contrast to the majority of the literature, we take a Bayesian perspective to estimation, and give emphasis to probabilistic prediction and real-time density forecasts.

More broadly, we contribute to an empirical literature that stresses the importance of modeling 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), Brunnermeier et al. (2020), and Carriero et al. (2022a). A distinct feature of our model is that we introduce fat tails as an independent (additive) outlier component, which can complement fat-tailed innovations or shocks in macroeconomic models.

[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 et al. (2016) stress the importance of accounting for the asynchronous nature of macroeconomic data within a DFM.

Our application to data from 2020 and thereafter relates to other attempts at assessing and improving macroeconometric models during that period, including 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. Ng (2021) uses pandemic related indicators as controls to clean the data prior to estimation of different econometric models. The 2020 episode is a useful case study of how heterogeneous dynamics and fat tails change the way that incoming data is interpreted by the DFM.[6]

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 our main empirical application, the out-of-sample evaluation exercise for 2000-2019, are presented in Section 4. Section 5 contains the additional application of our model to the spring 2020 recession and beyond. 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 data display (additive) 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. This captures

[6] Chetty et al. (2020) propose to use newly available high-frequency data to track the economy in real time. In an earlier version of this paper (Antolin-Diaz, Drechsel, and Petrella, 2021) we show how to integrate such time series into the DFM framework.

the fact that many time series related to real activity feature large one-off innovations to the level, such as strikes and weather related disturbances. These are purely transitory, 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, as visible in Panel (b) of Figure 1.[7]

Following Antolin-Diaz et al. (2017) we allow for low-frequency changes in the long-run growth rate of $\mathbf{y}_t$, which 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,

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

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.[8] We estimate the model on a panel of real activity variables and focus on the case of a single factor ($k = 1$, $\mathbf{f}_t = f_t$). We expand on the specific selection of real activity variables further below. The relevant laws of motion are

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

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 independent from the factor and idiosyncratic innovations and 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

[7] 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 is not applied to them.

[8] 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 the remaining variables are incorrectly specified as a constant. This is also true if the long-run components are shared with GDP.

of the model. The fat-tailed distribution is obtained by a Gaussian scale mixture ($o_{i,t} = \sqrt{\psi_{i,t}} z_{i,t}$ with $z_{i,t} \sim N(0, \sigma_{o,i}^2)$), where the scale mixture is a latent variable (Geweke, 1993; Jacquier et al., 2002).[9] Following Primiceri (2005), the time-varying parameters are driftless random walks:

$$a_t = a_{t-1} + v_{a,t}, \qquad v_{a,t} \overset{iid}{\sim} N(0, \omega_a^2) \tag{6}$$

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

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

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

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:

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

and where only one every third observation of x_t^q is observed. This reduces the presence of mixed frequencies to a problem of missing data in a monthly model.[10] Our Bayesian method exploits the state space representation of the DFM and jointly estimates the latent factors, the time-varying 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. Longer lags

[9] Specifically, assuming that $\psi_{i,t}$ is distributed i.i.d. inverse gamma, or that $v_i/\psi_{i,t} \sim \chi^2_{v_i}$, the marginal distribution of $o_{i,t} = \sqrt{\psi_{i,t}} z_{i,t} \sim t_{v_i}(0, \sigma_{o,i}^2)$.

[10] 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.

did not alter any results. 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, and to zero in subsequent lags. This prior reflects the belief that the factor is a cross-sectional average of the standardized variables. 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 $u_{i,t}$, a specification that is common in the literature (see, e.g., Banbura et al., 2011). 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.[11] For v_i, we use a weakly informative prior specified as a Gamma distribution $\Gamma(2, 10)$ (in line with Juárez and Steel, 2010) discretized on the support [3; 40]. The lower bound at 3 enforces the existence of a conditional variance.

2.4 Estimation algorithm

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. To avoid the use of large state-spaces we

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

propose a hierarchical Gibbs sampler, adapting ideas from Moench et al. (2013). The resulting algorithm allows parallelization of many steps, leading to extremely fast computation.

Let $\mathbf{\Phi} = \{\phi_j\}_{j=1}^p$, $\boldsymbol{\rho}_i = \{\rho_{i,j}\}_{j=1}^q$ and $\boldsymbol{\rho} = \{\boldsymbol{\rho}_i\}_{i=1}^n$ contain the autoregressive parameters for factor and idiosyncratic components, let $\boldsymbol{\lambda} = \{\{c_i, \{\lambda_{i,j}\}_{j=1}^m\}_{i=1}^n\}$ collect the constant terms and the loadings in the measurement equation, let $\boldsymbol{\omega}_\eta = \{\omega_{\eta,i}^2\}_{i=1}^n$ collect the variance of the SV of the idiosyncratic component, and let $\boldsymbol{\sigma}_o = \{\sigma_{o,i}^2\}_{i=1}^n$ and $\boldsymbol{\nu} = \{\nu_i\}_{i=1}^n$ denote the coefficients of the outlier components. We denote with $\boldsymbol{\theta}$ the vector containing all the parameters of the model, $\boldsymbol{\theta} \equiv \{\boldsymbol{\lambda}, \mathbf{\Phi}, \boldsymbol{\rho}, \omega_a^2, \omega_\varepsilon^2, \boldsymbol{\omega}_\eta, \boldsymbol{\sigma}_o, \boldsymbol{\nu}\}$ and $\boldsymbol{\theta}_{-j}$ denote the vector of parameters $\boldsymbol{\theta}$ with the exclusion of the element j. The latent components of the model are $\tilde{\boldsymbol{f}} = \{a_t, f_t\}_{t=1}^T$, $\boldsymbol{u}_i = \{u_{i,t}\}_{t=1}^T$ and $\boldsymbol{o}_i = \{o_{i,t}\}_{t=1}^T$. $\boldsymbol{\sigma}_\varepsilon = \{\log \sigma_{\epsilon,t}\}_{t=1}^T$, $\boldsymbol{\sigma}_\eta = \{\boldsymbol{\sigma}_{\eta,i}\}_{i=1}^n$ and $\boldsymbol{\sigma}_{i,\eta} = \{\log \sigma_{\eta_{i,t}}\}_{t=1}^T$ collect the stochastic volatilities of the model. The scale mixture required to generate the outlier components are collected into $\boldsymbol{\psi} = \{\boldsymbol{\psi}_i\}_{i=1}^n$ where $\boldsymbol{\psi}_i = \{\psi_{i,t}\}_{t=1}^T\}$. Let $\boldsymbol{y}$ collect the vector of data and let us also define the unobservable vector of outlier adjusted and interpolated variables (for the quarterly data) $\boldsymbol{y}_i^{OA} = \{\Delta y_{i,t}^{OA}\}_{t=1}^T$ where $\Delta y_{i,t}^{OA} = \Delta(y_{i,t} - o_{i,t}^j)$. Online Appendix A 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. *For each variable, $i = 1, \ldots, n$:*

   1.1. *Conditional on $\tilde{\boldsymbol{f}}$ and $\boldsymbol{\lambda}$, define $\tilde{\boldsymbol{y}}_i = \{\Delta y_{i,t} - c_{i,t} - \lambda_i(L) f_t\}_{t=1}^T$. Obtain draws from $p(\boldsymbol{y}_i^{OA}, \boldsymbol{u}_i, \boldsymbol{o}_i | \boldsymbol{y}, \tilde{\boldsymbol{f}}, \boldsymbol{\theta}, \boldsymbol{\sigma}_\varepsilon, \boldsymbol{\sigma}_\eta, \psi) = p(\boldsymbol{y}_i^{OA}, \boldsymbol{u}_i, \boldsymbol{o}_i | \tilde{\boldsymbol{y}}_i, \boldsymbol{\sigma}_{\eta_i}, \sigma_{o,i}^2, \boldsymbol{\psi}_i)$ using the Kalman filter and simulation smoother.*

   1.2. *Conditional on $\boldsymbol{o}_i$ obtain a draw from $p(\sigma_{o,i}^2 | \boldsymbol{o}_i, \boldsymbol{\psi}_i, \nu_i)$, $p(\boldsymbol{\psi}_i | \boldsymbol{o}_i, \sigma_{o,i}^2, \nu_i)$ and $p(\nu_i | \boldsymbol{o}_i, \boldsymbol{\psi}_i, \sigma_{o,i}^2)$ following Jacquier et al. (2004).*

2. *Conditional on the outlier adjusted and interpolated variables, $\boldsymbol{y}^{OA} = \{\boldsymbol{y}_i^{OA}\}_{i=1}^n$:*

   2.1. *Draw from $p(\tilde{\boldsymbol{f}} | \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}_\varepsilon, \boldsymbol{\sigma}_\eta, \psi) = p(\tilde{\boldsymbol{f}} | \tilde{\boldsymbol{y}}, \boldsymbol{\lambda}, \mathbf{\Phi}, \omega_a^2, \boldsymbol{\sigma}_\varepsilon, \boldsymbol{\sigma}_\eta)$ using the Kalman filter and simulation smoother on the outlier adjusted and interpolated data, where the $\{i, t\}$-th element of $\tilde{\boldsymbol{y}}$ is defined the quasi-difference of the outlier adjusted indicator $y_{i,t}^{AO} - \sum_{j=1}^q \rho_{i,j} y_{i,t-j}^{AO}$.*

   2.2. *Conditional on $\tilde{\boldsymbol{f}}$ and given the standard normal and inverse-gamma priors draw from $p(\omega_a^2 | \boldsymbol{a})$, $p(\mathbf{\Phi} | \boldsymbol{f}, \boldsymbol{\sigma}_\varepsilon)$ and $p(\boldsymbol{\lambda} | \tilde{\boldsymbol{y}}, \tilde{f}, \boldsymbol{\sigma}_\eta)$ are easily obtained from the inverse-gamma and*

Normal distribution and Draws from $p(\boldsymbol{\sigma}_\varepsilon|\tilde{f}, \boldsymbol{\Phi})$ and $p(\omega_\varepsilon^2|\boldsymbol{\sigma}_\varepsilon)$ can be obtained following Kim et al. (1998).

2.3. *Conditional on $\tilde{\boldsymbol{f}}$ and $\boldsymbol{\lambda}$, for each variable $i = 1, ..., n$ define the t-th element of $\tilde{\mathbf{u}}_i$ as $u_{i,t} = y_{i,t}^{AO} - c_{i,t} - \sum_{j=1}^{m} \lambda_{i,j} f_{t-j}$, draws of $p(\boldsymbol{\rho}_i|\tilde{\mathbf{u}}_i, \boldsymbol{\sigma}_\eta)$ can be obtained from the Normal distribution, and draws from $p(\boldsymbol{\sigma}_{i,\eta}|\tilde{\mathbf{u}}_i, \boldsymbol{\rho})$ and $p(\omega_{i,\eta}^2|\boldsymbol{\sigma}_{i,\eta})$ can be obtained following Kim et al. (1998)*

3. *Go back to Step 1 until convergence has been achieved.*

We note several points. First, the algorithm iterates between a univariate state space in Step 1, which performs outlier adjustment, and a multivariate one in Step 2, 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 1 is independent across variables, so it can be run in parallel using multi-core processors. Third, exploiting a quasi-difference representation of the data (see Kim and Nelson, 1999 or Bai and Wang, 2015), one can avoid to include the idiosyncratic component among the state vector, which leads to a substantial reduction in the dimensionality of the state-space system in Step 2.

2.4.1 Efficient precision sampler with missing observations

In order to make the algorithm more efficient we use the vectorized version of the Kalman filter/smoother (sometimes referred to as precision sampler, see Chan and Jeliazkov, 2009). The vectorized approach leads to substantial gains in computational time.

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

$$
\begin{aligned}
\mathbf{y} &= \overline{\mathbf{H}}\mathbf{x} + \eta, \qquad \eta \sim N(0, \overline{\mathbf{R}}) \\
\overline{\mathbf{F}}\mathbf{x} &= \mathbf{x}^* + \mathbf{e}, \qquad \mathbf{e} \sim N(0, \overline{\mathbf{Q}})
\end{aligned}
\tag{10}
$$

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's matrices are organized as

$$\overline{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_1 & & & \\ & \mathbf{H}_2 & & \\ & & \ddots & \\ & & & \mathbf{H}_T \end{bmatrix}, \quad \overline{\mathbf{R}} = \begin{bmatrix} \mathbf{R}_1 & & & \\ & \mathbf{R}_2 & & \\ & & \ddots & \\ & & & \mathbf{R}_T \end{bmatrix},$$
$$\overline{\mathbf{F}} = \begin{bmatrix} \mathbf{I}_k & & & \\ -\mathbf{F}_1 & \mathbf{I}_k & & \\ & \ddots & \ddots & \\ & & -\mathbf{F}_{T-1} & \mathbf{I}_k \end{bmatrix}, \quad \overline{\mathbf{Q}} = \begin{bmatrix} \mathbf{P}_0 & & & \\ & \mathbf{Q}_1 & & \\ & & \ddots & \\ & & & \mathbf{Q}_{T-1} \end{bmatrix}. \tag{11}$$

In Proposition 1 we extend the results of Chan and Jeliazkov (2009) to the case where there are missing observations.

Proposition 1. *Let $\mathbf{J}$ be the diagonal selection matrix selecting the full rank portion of the generic time t covariance matrix of the transition equation and $\mathbf{\Xi}$ a diagonal matrix with ones corresponding to the existing elements in $\mathbf{y}$ and zero for the missing elements. Define $\overline{\mathbf{J}}$=$(\mathbf{I}_T \otimes \mathbf{J})$, $\widetilde{\mathbf{F}} = \overline{\mathbf{J}}'\overline{\mathbf{F}\mathbf{J}}$, $\widetilde{\mathbf{Q}} = \overline{\mathbf{J}}'\overline{\mathbf{Q}\mathbf{J}}$, $\widetilde{\mathbf{H}} = \overline{\mathbf{H}\mathbf{J}}$, and $\widetilde{\mathbf{R}}^{-1} = \mathbf{\Xi}'\overline{\mathbf{R}}^{-1}\mathbf{\Xi}$. The state vector $\mathbf{x} \sim \mathbf{N}(\varkappa, \mathbf{P})$ with*

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

where $\mathbf{K} = \widetilde{\mathbf{F}}'\widetilde{\mathbf{Q}}^{-1}\widetilde{\mathbf{F}}$.

The precision algorithms take advantage of the sparse and banded structure of many of the very large matrices that enter into the posterior for the states. The introduction of $\mathbf{J}$ allows us to deal with the presence of a rank deficient system for the state variables, arising from the presence of multiple lags in the dynamics of the factors,[12] whereas $\mathbf{\Xi}$ deals with the missing observations.[13]

[12] $\mathbf{J}$ is a k-dimensional diagonal matrix with zeros corresponding to singular columns and one for the nonsingular columns of $\mathbf{Q}_t$, where the latter denotes the covariance matrix of the transition equation innovations of the (original) state space model.

[13] In MATLAB, the use of the 'backslash' operator as opposed to the standard inverse operator guarantees further efficiency in computation. Given a $k \times k$ non-singular sparse matrix $\mathcal{S}$ and a $k \times 1$ vector x, we have that $\mathcal{S}^{-1}x \equiv \mathcal{S}\backslash x$, which denotes the unique solution for z to the system $\mathcal{S}z = x$. Defining the Cholesky factor $\mathcal{C}$, such that $\mathcal{C}\mathcal{C}' = \mathcal{S}$,$\mathcal{S}^{-1}x = \mathcal{C}'\backslash(\mathcal{C}\backslash x)$, and this solves two triangular systems by forward substitution followed by back substitution.

Table 1: DATA SERIES USED FOR 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.

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. This follow insights from the DFM literature. 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. Specifically, we include all of the available surveys of consumer and business sentiment, because the literature has highlighted how 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. As for the possible inclusion of disaggregated series within each category, Boivin and Ng (2006), Banbura et al. (2013), and Alvarez et al. (2012) argue that the presence of strong correlation in the idiosyncratic components of disaggregated series of the same category will be a source of misspecification that can worsen the performance of the model in terms of in-sample fit and out-of-sample forecasting of key series. Therefore, we use a medium-sized panel of 29 series with representative indicators of each category, which are listed with detailed information in Table 1.

3 General implications for nowcasting economic activity

In this section, we show how the explicit modeling of heterogeneous dynamics and fat tails fundamentally modifies the way that a DFM interprets incoming information in real time. This analysis is based on estimating our Bayesian DFM on data from 1947 to 2019. To set the stage, we introduce some relevant definitions. In any filtering problem, prediction errors for the observables map into updates of the estimate of the latent states (Harvey, 1989). 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"

$$E(f_{t_k}|\Omega_2) - E(f_{t_k}|\Omega_1) = w_{j,t}\left(y_{j,t_j} - E(y_{j,t_j}|\Omega_1)\right), \tag{12}$$

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. The weight $w_{j,t}$ is the slope of the influence function.[14]

[14]When more than one variable is released, equation (12) can be used to obtain a "news decomposition", 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

The addition of heterogeneous dynamics and fat tails will affect the influence function in different ways. The presence of heterogeneous dynamics means that the conditional expectation $E(y_{j,t_j}|\Omega_1)$ in the standard model is misspecified by the omission of lags of the factor in the information set, whereas the presence of fat tails means that the slope $w_{j,t}$ of the influence function, which is assumed to be linear in the standard model, is in fact non-linear and non-monotonic. The analysis in this section will reveal that both of these sources of misspecification lead the standard DFM to discard information contained in the "hard" data series. We also show that lead-lag dynamics and fat tails meaningfully interact with each other.

3.1 Heterogeneous lead-lag dynamics

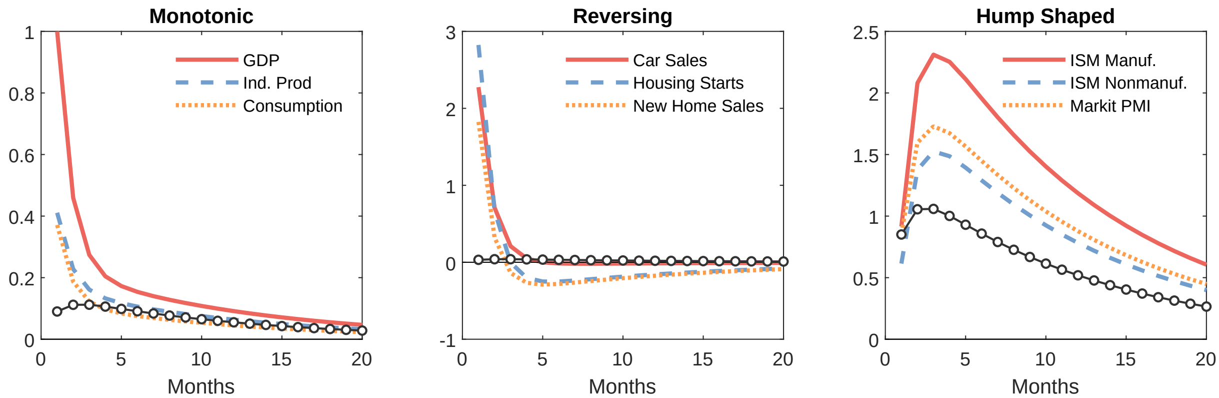
In a standard DFM, without lags of the factor included in the measurement equation ($m = 0$), the conditional expectation of the future value of an individual series is proportional to the estimate of the common factor, i.e. $E(y_{j,t_j}|\Omega_1) = \lambda_j E(f_t|\Omega_1)$. Equivalently, the impulse response function (IRF), which measures the revision of this conditional expectation for current and future periods after an innovation to the common factor, can be written $\frac{\partial y_{j,t+h}}{\partial \varepsilon_t} = \lambda_i \frac{\partial f_{t+h}}{\partial \varepsilon_t}$, so it inherits the dynamics of the IRF of the factor itself. 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).[15] The omission of relevant lags in the conditional expectation function for each variable will lead to inefficient and volatile forecast errors in Equation (12), and therefore to erratic updates to the GDP nowcast. Moreover, omitting lags may bias the estimated factor itself towards over weighting certain groups of variables and under weighting others.

Figure 2 plots the IRFs of different variables in our estimated model and groups them according to their estimated shapes. To visualize the differences with a standard DFM, in each panel we superimpose the average IRFs for the same variables from a standard DFM without heterogeneous dynamics and without fat tails (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). A number of other

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).

[15] 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 2.

Figure 2: IRFS OF SELECTED VARIABLES TO AN INNOVATION IN THE COMMON FACTOR

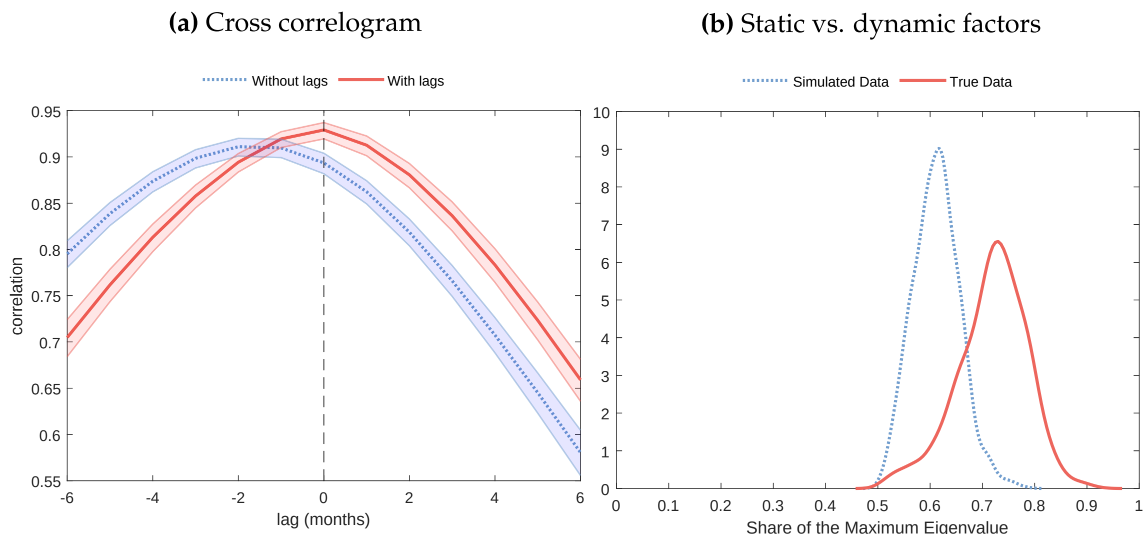


Notes. IRFs of different variables to an innovation in the process of the dynamic factor (an innovation to ε in Equation (3)). In each panel, the black line with circular markers shows the average across IRFs with homogeneous dynamics ($m = 0$), while the solid red, orange dotted and dashed blue lines 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').

variables, particularly those related to investment and durable spending, 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 hump-shaped dynamics (right panel).

Across the three panels, it is visible that the IRFs in the basic DFM without lead-lag dynamics inherit the hump 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 inherit their properties. This is unappealing when data series that mean-revert more quickly may provide a meaningful signal of economic activity not captured by the surveys. Moreover, the close to zero IRF in the basic DFM associated with the variables that tend to lead the cycle (the group of "reversing" variables) is evidence that in the standard model their estimated loading is close to zero. As a consequence of their

Figure 3: UNPACKING HETEROGENEOUS DYNAMICS



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).

asynchronous relationship with the business cycle the standard model perceives them as largely uncorrelated with the common factor and discards any useful information contained in them.[16]

To analyze the problem of ignoring lead-lag dynamics further, Figure 3, Panel (a) shows that in a specification without 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. In a basic DFM the common factor mainly captures the information in the soft variables, and is therefore not able to capture the higher frequency variation of key macro indicators such as industrial production or consumption. As a consequence the estimated factor lags GDP growth, which implies that the nowcasts are inheriting this 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.

[16] For instance, in the standard model, the common factor, which captures roughly 80% of the dynamics of GDP growth and almost 90% of the variation in GDI, explains up to 10% of the overall month-on-month variation of initial claims (and less 15% of the year-on-year changes).

3.1.1 Relation to alternative specifications with multiple factors

An implication of our modeling choice of one factor ($k = 1$) and heterogeneous dynamics ($m > 0$) is 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. In the case of the latter paper, the additional factors load only on subgroups of variables and their motivation is to capture clusters of correlation in the idiosyncratic components rather than dynamic heterogeneity in response to the common factor. These alternative DFM specifications are related to the one we propose, 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.[17] Which specification is preferred amounts to asking how close to reduced rank is the covariance matrix of a multiple factor specification in the data. To test this, we estimate a homogeneous DFM with two static factors for our data set, and compute, for each posterior draw, the share of the largest eigenvalue in the factor covariance matrix. The result, displayed in panel (b) of Figure 3, 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, which leads us to conclude that the data provide stronger support for the specification with one factor and lead-lag dynamics. More broadly, these results support the idea that a single aggregate innovation, transmitted heterogeneously across variables, indeed captures common fluctuations in real activity variables.[18]

3.2 Fat-tailed outliers

The explicit modeling of fat-tailed outliers affects the slope $w_{j,t}$ of the influence function in Equation 12. In a linear Gaussian state space model this slope corresponds to the j-th column of

[17] 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).

[18] This innovation can span multiple *structural* macroeconomic shocks and innovations to the common factor identify the combination of shocks that best capture in all data in line with Angeletos et al. (2020). Identifying these structural shocks separately would require additional variables, such as for instance prices, as well as additional assumptions.

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

$$w_{j,t}(y_{j,t_j}) = \frac{E\left((f_{t_k} - f_{t_k|\Omega_1})(f_{t_j} - f_{t_j|\Omega_1})\right)\Lambda_j'}{\Lambda_j E\left((f_{t_j} - f_{t_j|\Omega_1})(f_{t_j} - f_{t_j|\Omega_1})\right)\Lambda_j' + \sigma^2_{\eta_{j,t_j}} + \psi_{j,t_j}\sigma^2_{o,j}}. \tag{13}$$

The expectation terms $E\left((f_{t_k} - f_{t_k|\Omega})(f_{t_j} - f_{t_j|\Omega})\right)$ in (13) denote the filtering uncertainty about the common factor. Notice first that since our model also features SV, 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.[19] On the contrary, when the SV 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. In addition, Equation (13) makes clear that in our model $w_{j,t}$ depends on the data itself. The reason is the term ψ_{j,t_j}, which takes the form

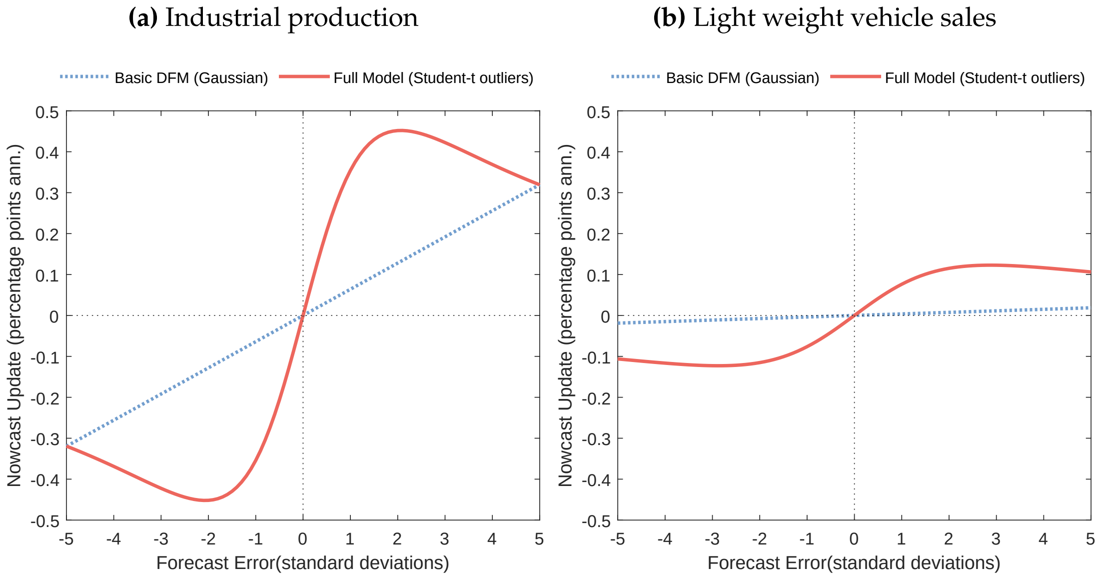
$$\psi_{j,t_j} = (((y_{j,t_j} - E(y_{j,t_j}|\Omega_1))^2/\sigma^2_{o,j} + v_{o,j})/(v_{o,j} + 1), \tag{14}$$

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 4 plots the influence functions for two example variables, industrial production and car sales, both for the linear case without outliers (dotted blue) and the nonlinear case

[19] 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 SV, in which uncertainty always declines in response to new data, as discussed, e.g., by Banbura and Modugno (2014).

Figure 4: INFLUENCE FUNCTIONS IN DFMS WITHOUT AND WITHOUT FAT TAILS

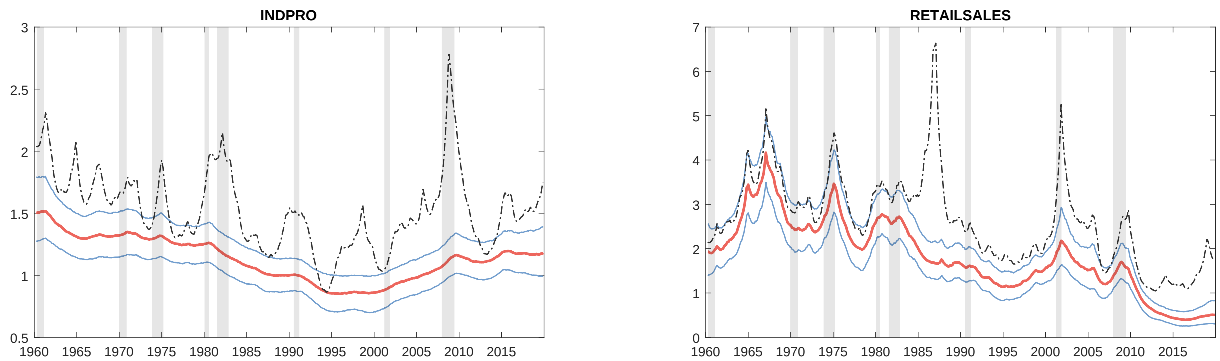


Notes. Influence functions for industrial production and light weight vehicle sales. An influence function indicates by how much the estimate of the common factor is updated when the release in the variable is different from its forecast and thus contains "news." The dotted blue line plot these influence functions in the Gaussian case, while the red lines represent the Student-t case. As shown in equations (13)-(14) 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.

with outliers (solid red). The influence functions for all variables are presented in the Online Appendix. For a model with Gaussian innovations, the update of the factor is a line with slope equal to $w_{j,t}$. Noisy variables such as car sales have very low weights, meaning that surprises to this variable lead to small updates to the current factor. With t-distributed outliers, the influence functions are now S-shaped. Around the origin, 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 that 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 removed.

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 the "soft" business surveys. This implies an increase in the slope of their influence functions. At the same time, the hard data series are

Figure 5: STOCHASTIC VOLATILITY ESTIMATES WITH AND WITHOUT OUTLIERS

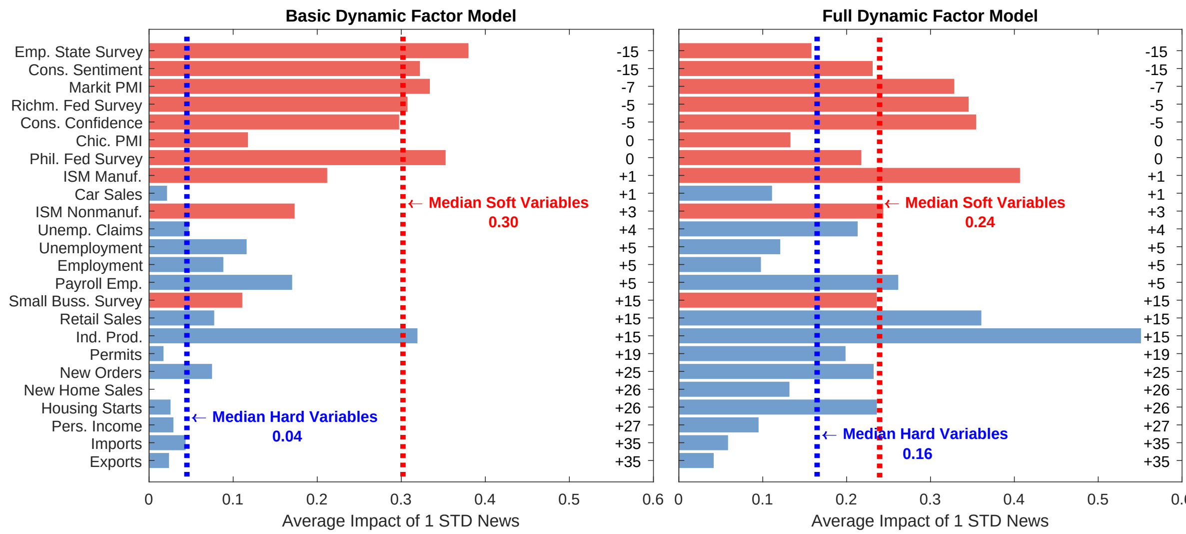


Notes. For each of the two panels, the broken black line reports the posterior median estimate of the stochastic volatility processes associated with each of the two indicators using a model with SV but without outliers. The solid red and blue lines are, respectively the median and 68% HPD intervals in the full model with SV and with outliers. Shaded areas indicate NBER recessions.

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 stochastic volatility. A DFM with SV features a time-varying Kalman gain. Therefore the weight given to the "news" to each of the variables is time-varying as a reflection of relative volatility of the common factor and the idiosyncratic component (specifically, this is equal to the expression in Equation (13) with $\sigma_{o,j} = 0$). A large idiosyncratic outlier occurring at time t will lead to large revisions in the estimate of the SV associated with it, and hence, a temporary reduction of the weight of news to that indicator in the estimate of the factor. However, since the SV is specified as a persistent process, this will lead to a protracted increase of the SV, meaning that a one-off outlier leads to discounting of the information content in the variable for many consecutive months. In Figure 5 we report the estimated SV for two key indicators of economic activity: industrial production and retail sales for a model with and without outlier component. The broken black line reports the posterior median estimate of the SV processes associated with each of the two indicators using a model with SV but without outliers, as in Antolin-Diaz et al. (2017).

Figure 6: CONTRIBUTION OF NEWS IN ECONOMIC INDICATORS TO GDP GROWTH NOWCASTS



Notes. Impact of news in a variable on the update of the GDP nowcast, in a basic DFM (left panel) and the Bayesian DFM with lead-lag dynamics and fat tails (right panel). Red bars represent soft indicators, blue bars represent hard indicators. The numbers associated with each indicator represent the release lag in days relative to the end of the reference period. In the model with fat tails we approximate the slope of the influence function around the origin as the slope of the line between a +/- 0.5-standard deviation forecast error. The influence function is in general time-varying and we report the average of its slope in a representative year.

The solid red and blue lines are, respectively the median and 68% HPD intervals in the full model with outliers. The figure shows that the model with SV only features cyclical variation in the idiosyncratic volatilities, which appears inconsistent with their assumed random walk specification. Once outliers are taken into account, the SV process becomes more persistent, capturing only low-frequency developments in the idiosyncratic volatilities.

3.3 Implications for signal extraction from hard vs. soft data

Figure 6 compares the slope of the influence function, i.e the impact of typical news about a given variable on the GDP nowcast, across all variables in our panel. The figure presents this for standard DFM without heterogeneous dynamics or fat tails (left panel) and our full Bayesian DFM (right panel). Red bars represent soft indicators, while blue bars represent hard indicators. The numbers associated with each indicator represent the release lag in days relative to the end of the reference period, according to which the series have been ordered in the Figure.[20] The

[20]Both Figure 4 and 6 have been computed taking into account the usual pattern of missing data present on the day of the release of each series, so the influence functions fully reflect the impact of the timeliness of the releases.

left panel reveals that there is a stark difference in the impact of hard versus soft indicators. In fact, with the exception of industrial production, hard indicators have a very limited average impact on the GDP nowcasts, irrespective of their timeliness. For instance, relatively timely hard indicators such as car sales or unemployment claims, receive very little weight. The median impact of the soft variables is more than 7 times higher than that of the hard variables. The right panel shows how in our full model, hard macroeconomic indicators contribute to the nowcast updates almost as much as soft indicators, with a median impact of 0.16 relative to 0.24. Even data that is released with a significant lag, such as housing starts, still provide a signal about activity that is incorporated in the model's GDP nowcasts. Our findings contradict the conclusion of the literature that soft data receives more weight in the nowcasting process due only to its timeliness (Bańbura and Rünstler, 2011; Banbura et al., 2011, 2013), and point to misspecification of the model as an additional reason.

The analysis above has established that explicitly modeling of heterogeneous dynamics and fat tails substantially changes the interpretation of macroeconomic data in a DFM for the purposes of nowcasting. The next section evaluates whether and how these novel features of our Bayesian DFM translate into improved out-of-sample performance.

4 Real-time out-of-sample model evaluation

This section presents our main empirical application. We estimate the model using data going back to 1947, construct daily estimates of US GDP growth from 2000 to 2019, and formally evaluate the point and density forecasting performance of our model relative to benchmark competitors. 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 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 intensive. It is made feasible thanks to our efficient algorithm and by exploiting modern cloud computing infrastructure. Details are provided in Online Appendix C.

4.1 Model forecasts and GDP releases over time

Figure 7, Panel (a) 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 Bayesian DFM. As the data arrive, the model tracks the developments of the last two decades in real time, such as the early 2000s recession and the Great Recession in 2008-09.[21]

Panel (b) compares 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 estimated on the same data set but without 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 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. 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.

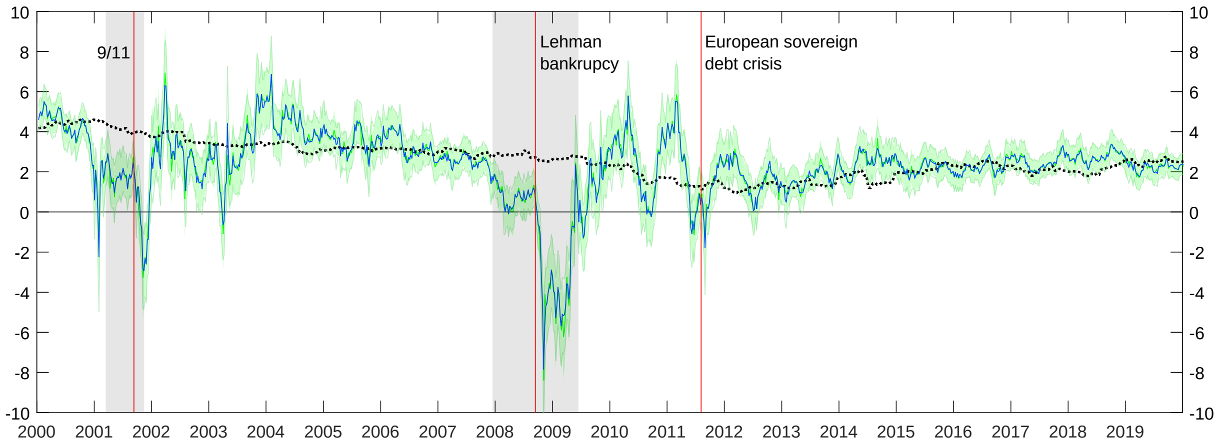
The shaded areas in the figure represent a 68% HPD interval for each model. Thus, if the density forecasts were well calibrated, out of the 80 quarterly observations we would expect about 26 of them, or 32%, to fall outside of the bands. For the case of the basic model, 17 or 21% fall outside, whereas for the full model it is 27, or 34%.[22] This indicates that the basic model,

[21]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).

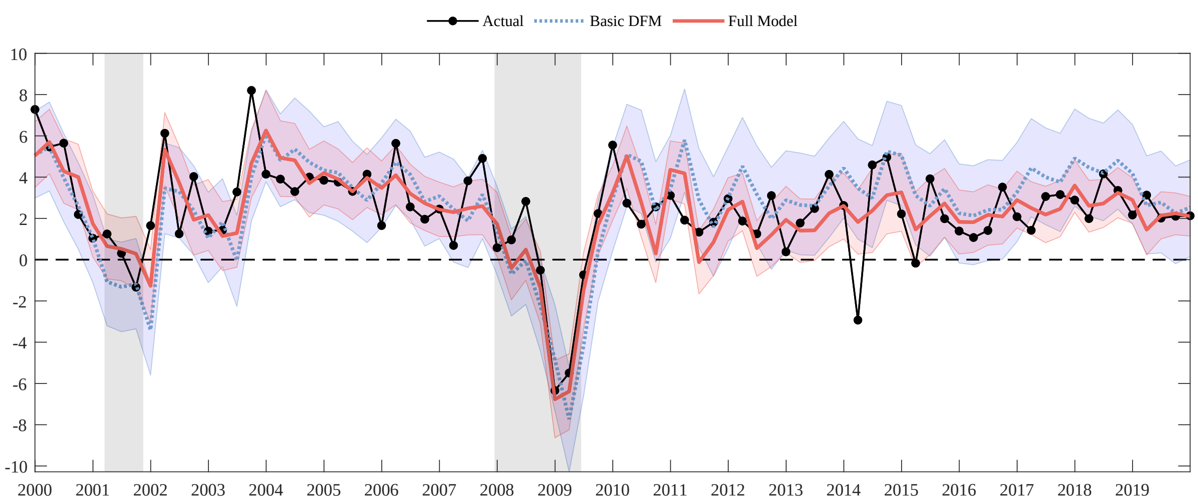
[22]Applying a formal test for the correct calibration of the density forecast (Rossi and Sekhposyan, 2019) rejects (at 5% level) the correct specification for the basic DFM model, but does not reject when applied to the forecasts of the full model. This is true when the test is applied to the entire density, as well as when this is applied to the left and right side of the density separately.

Figure 7: Real-time nowcasts over the 2000-2019 period

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



(b) Realized GDP growth and alternative nowcasts



Notes. Panel (a) plots the daily estimate of US real GDP growth (blue line) with associated uncertainty (green shaded areas indicates the 68% posterior credible intervals) computed from our Bayesian DFM from 2000 through to the end of 2019. The vertical red lines indicate significant events. Panel (b) shows, as the black line, the third release of real GDP growth 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 68% posterior credible intervals around around the nowcasts. In both panels, the vertical gray shaded areas indicate NBER recessions.

which does not feature time-varying volatility, is on average too conservative, producing bands that are too wide. The restrictive assumption of constant volatility is behind this result: the sample features long, stable expansions punctuated by relatively large recessions, so a single, average, estimate of volatility is an overestimation most of the time.

4.2 Real-time evaluation and comparison to alternative models

Table 2 provides a formal forecast evaluation of our model against a basic DFM, the model with time-varying long-run growth and SV proposed in Antolin-Diaz et al. (2017), and the DFM maintained by the New York Fed staff, described in Bok et al. (2018).[23] The table formally evaluates the point and density forecast accuracy relative to the third release of GDP.[24]

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 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 *forecasts* 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 all models become more accurate as the forecasting horizon shrinks and more information comes in. Importantly, our full Bayesian DFM produces significantly better point forecasts than the basic DFM at all horizons. Moreover, the our model also outperforms the model proposed in Antolin-Diaz et al. (2017), which features time-varying long run growth and SV but not heterogeneous dynamics nor fat tails, as well as the NY Fed Model, with economically large improvements. At the 45-day horizon, the RMSE from our model is 17% lower than the NY Fed model, and highly statistically significant. The differences decline by the time of the first release of GDP, 30 days after the reference quarter.

Panel (b) compares 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

[23]The New York Fed Staff Nowcast was updated weekly over the period we analyze. Historical nowcasts are available online, which allows a direct comparison. See https://www.newyorkfed.org/research/policy/nowcast.html.

[24]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 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 unchanged, but we find that all models do generally better at forecasting less "mature" vintages. If the objective was 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 2: OUT-OF-SAMPLE PERFORMANCE OF DIFFERENT MODELS

(a) Point forecasting: RMSE	Full Model	ADP 2017	Basic DFM	NY Fed
–180 days	2.26	2.40	2.53	–
		[0.16]	[0.02]	–
–90 days (start reference quarter)	2.14	2.42	2.57	2.36
		[0.09]	[0.01]	[0.29]
–60 days	1.88	2.04	2.21	2.18
		[0.05]	[0.01]	[0.05]
–45 days (middle reference quarter)	1.66	2.03	2.09	1.98
		[0.05]	[0.00]	[0.02]
–30 days	1.66	1.92	2.09	1.83
		[0.05]	[0.00]	[0.07]
0 days (end reference quarter)	1.48	1.83	1.98	1.67
		[0.07]	[0.00]	[0.08]
+30 days (first release)	1.48	1.79	1.96	1.60
		[0.03]	[0.00]	[0.13]

(b) Density Forecasting: Log Score	Full Model	ADP 2017	Basic DFM	NY Fed
–180 days	–2.16	–2.16	–2.40	–
		[0.00]	[0.00]	–
–90 days (start reference quarter)	–2.09	–2.21	–2.37	–2.28
		[0.00]	[0.01]	[0.05]
–60 days	–1.98	–2.05	–2.24	–2.17
		[0.00]	[0.00]	[0.01]
–45 days (middle reference quarter)	–1.89	–2.07	–2.18	–2.08
		[0.00]	[0.04]	[0.00]
–30 days	–1.88	–2.00	–2.17	–1.99
		[0.00]	[0.00]	[0.00]
0 days (end reference quarter)	–1.81	–1.98	–2.14	–1.91
		[0.00]	[0.00]	[0.00]
+30 days (first release)	–1.80	–1.95	–2.13	–1.87
		[0.00]	[0.00]	[0.00]

Notes. Comparison of the forecasting performance of different models: the full Bayesian DFM; the DFM model with stochastic trend and volatility (Antolin-Diaz et al., 2017, ADP 2017); the basic DFM; the New York Fed Staff Nowcast. We report the RMSE (top panel) and Log score (bottom panel) across various forecasting horizons. For the first three columns 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 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).

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 that the relative density forecasting performance of our Bayesian DFM is even stronger than the point forecasting performance. At 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.

In Online Appendix D we show that similar gains are found using alternative evaluation metrics for (mean absolute error and continuous rank probability score). We also evaluate the models over a continuous forecast horizon (rather than at a selected number of horizons) and show that while both models improve their performance as the information arrives over the nowcasting horizon, the gains from the full model are large throughout the forecast, nowcast and backcast horizons. Furthermore, we evaluate recursively the performance of the full model against the basic DFM. This exercise shows that the improvements from the full model are large and significant already after 2 years in the evaluation sample. Finally, the Online Appendix also demonstrates how our model can be used to asses tail risks in real time.[25]

4.3 Daily intraquarter evaluation of nowcasting performance

In practical applications, nowcasts are usually monitored at the daily frequency with the release of new information. Nowcast updates within the quarter can be considered as fixed-event forecasts and this provides another angle to evaluate a nowcasting model: efficient forecasting requires that forecast revisions should be uncorrelated with past revisions (Nordhaus, 1987).

Let x_t be the target variable observed at lower frequency, in our case GDP growth, we denote the daily forecast revision $r_{t,d} = E_{t-d}(x_t) - E_{t-d-1}(x_t)$, where d denotes the number of days to the end of the reference quarter. A simple test of forecast efficiency requires that revisions are not predictable using lagged revisions. We test this hypothesis pooling all the quarters in our

[25] See also Carriero et al. (2022a) for an alternative approach to nowcasting tail risks.

Table 3: OUT-OF-SAMPLE REVISIONS ANALYSIS

	AR(1) Coefficient	Variance
Full Model	-0.002	0.052
	[0.878]	
ADP 2017	-0.201	0.085
	[0.000]	
Basic DFM	-0.127	0.063
	[0.000]	

Notes. The left column reports the AR(1) coefficient revision $r_{t,d} = E_{t-d}(x_t) - E_{t-d-1}(x_t)$, for any day $d = [90, ..., -30]$ for which there is a release of a new data. The number in parenthesis denote the p-value associated to the null hypothesis of a zero autoregressive coefficient. The right panel reports the variance of the revisions.

out-of-sample period for at the nowcasting and backcasting window, $d = [90, ..., -30]$.

Table 3 reports the estimated autoregressive coefficients for the full model, the model of Antolin-Diaz et al. (2017), as well as the basic DFM. The latter two models display a significantly negative coefficient. In other words, both models display signs of 'overreaction' to the news in the data, with nowcast updates that are on average partially reversed with the subsequent release of new information. For the full model we cannot reject the null hypothesis of unpredictability. The same table reports a measure of the variance of the revisions, highlighting that the baseline is also the one with the smallest variability in the revisions. Heterogeneous dynamics and fat tails thus also contribute to stable and efficient daily updates of the nowcast.

4.4 Comparison to Greenbook and survey expectations

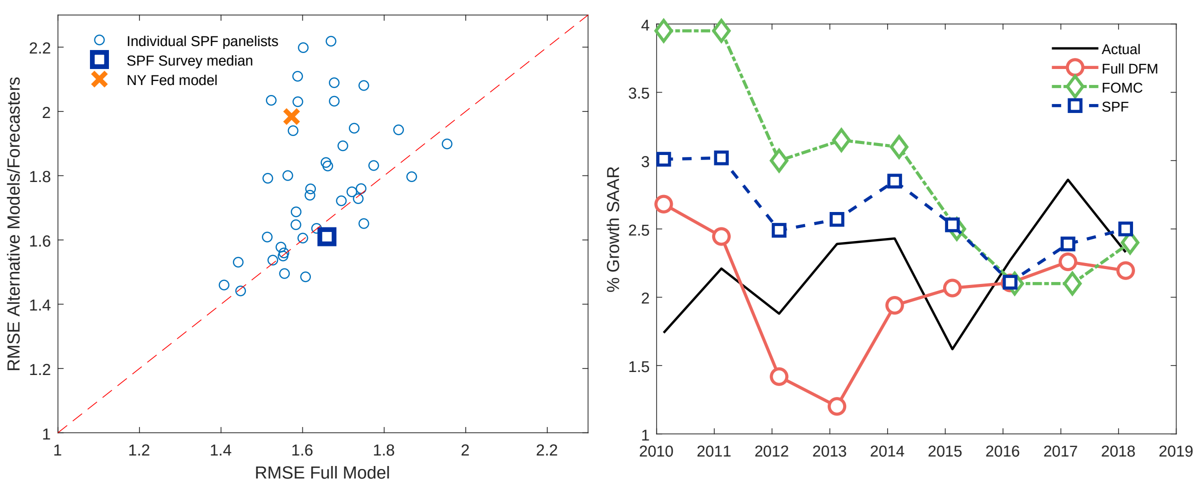
In addition to assessing the performance of our Bayesian DFM relative to other formal econometric models, we compare its forecasts to those of the Survey of Professional Forecasters (SPF) and its individual participants, as well as to the Federal Reserve Staff's Greenbook forecasts produced for the Federal Open Market Committee (FOMC). A comparison 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 8 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. In this panel, 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.[26] 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.

Figure 8: 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.

We also compare the quarterly GDP nowcasts of our model against the Federal Reserve

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

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 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.[27] 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.[28] Yet, in line with earlier findings of Romer and Romer (2000) and Nakamura and Steinsson (2018), Federal Reserve Staff forecasts continue to be a tough benchmark for formal models, appearing to possess considerable additional information in particular close to the end of the reference quarter.

Panel (b) considers longer-horizon forecasts, and plots 2-year ahead GDP growth 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 FOMC 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. In fact, 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. Moreover, 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.

[27] 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 1.82 vs. 1.55 of the GB (p-value $= 0$)

[28] 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.

5 Application to macro data during and after spring 2020

The scale and speed of the US recession of spring 2020 poses unique challenges for macroeconometric models. The New York Fed's nowcasting model, analyzed as a benchmark model in the previous section, was entirely suspended.[29] A nascent academic literature has studied these new challenges, mostly in the context of VAR models (Lenza and Primiceri, 2020; Schorfheide and Song, 2020; Cimadomo et al., 2020; Ng, 2021; Carriero et al., 2022b). In this section, we examine the behavior of our Bayesian DFM during this exceptional period. We focus mainly on our model as we found that for a basic DFM with Gaussian disturbances and constant volatility, the extreme nature of the 2020 observations would require us to adopt some ad-hoc procedure to censor outliers. We found that even the model with SV (but no outliers) is subject to enormous instability over this period. We begin by using the second quarter of 2020 as a case study of how the features of our model allow us to deal with the real-time data flow at the height of the pandemic. We then explore the evolution of real-time GDP nowcasts of the model over the full 2020-22 sample and compare it to the BEA's GDP releases. This highlights the strengths and weaknesses of our modeling innovations during this period.

5.1 Case Study: real-time assessment of economic activity in 2020:Q2

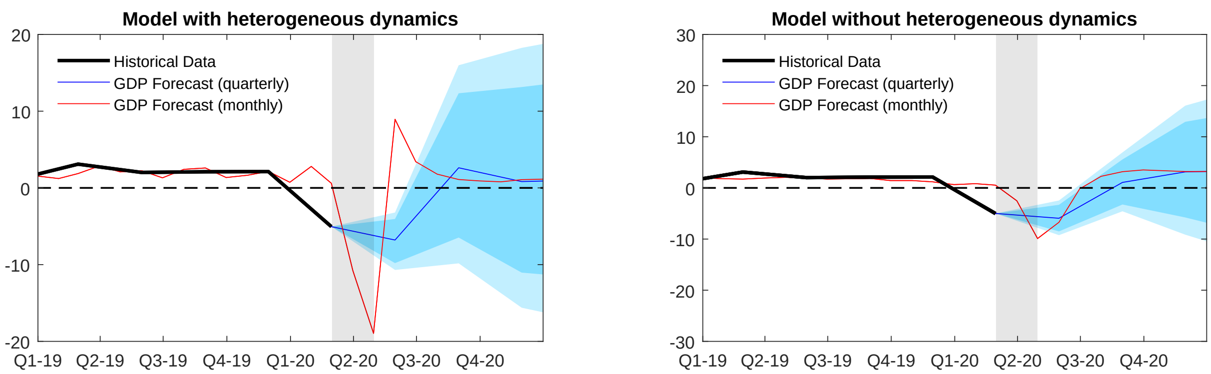
Figure 9 examines the highly uncertain first months of the pandemic, before any official releases of US GDP. Panel (a) reports the estimation results using data available on June 1, 2020. In the left panel, the full model is used, whereas in the right panel, the heterogeneous dynamics are switched off (i.e., $m = 0$). In both cases, we report the posterior mean of (latent) monthly GDP growth, the common component, and their forecasts through 2021. As discussed in Section 3.1, in the standard DFM the soft data dominate the estimation of the factor, so 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 the model to extract more information from hard data, which display much less persistence. As of June 1, soft data available for May already displayed a partial recovery, and some of the latest available hard data started to show the first

[29]This was communicated to the public in a suspension notification on September 3, 2021 on the New York Fed's website.

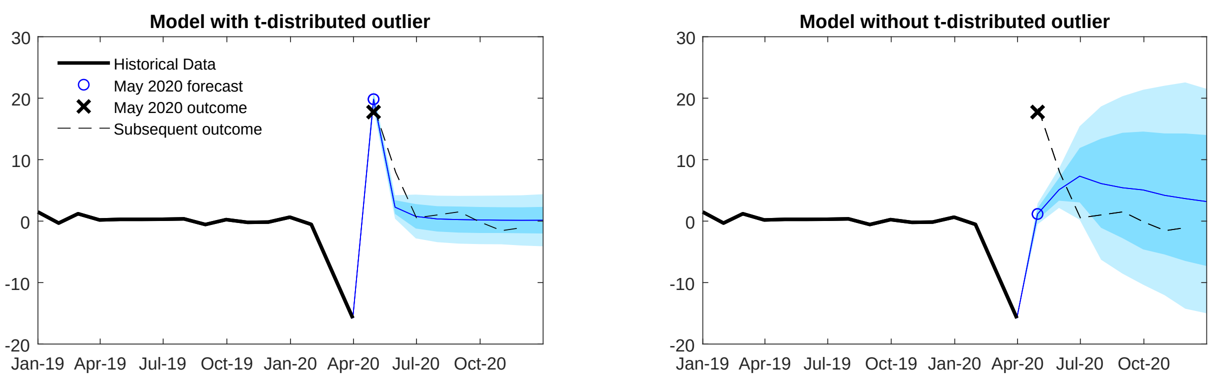
signs of an early recovery. For the with heterogeneous dynamics (left panel) this information translates into a forecast of a sudden large drop in activity followed by a quick, albeit partial, *rebound* in GDP. For the model with homogeneous dynamics (right panel) the forecast is instead for a shallower but more prolonged recession, similar to the 2008-2009 financial crisis.

Figure 9: 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. Shades denote the 68% and 90% posterior credible intervals.

Panel (b) of Figure 9 illustrates the role of the t-distributed outliers by looking at the model's real-time prediction of retail sales. This is one of the hard data series from which our Bayesian DFM extracts a much stronger signal than the standard DFM, as shown in Figure 6. 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.[30] The model interprets a large fraction of the April 2020 drop as one such outlier, and forecasts a strong rebound in May. Consistent with the transitory nature of this pattern, the idiosyncratic stochastic volatility remains at normal levels for the subsequent forecast horizon. The right panel reports the same forecast when the t-distributed outliers are switched off. As 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 the data 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.

5.2 GDP nowcasts during the period 2020-22

Figure 10 presents the evolution of GDP nowcasts produced by the model through to the middle of 2022, in comparison to the actual GDP releases. Panel (a) contains the daily estimates of US real GDP growth with information up to a given day, as well as the uncertainty associated with these estimates. The evolution of these daily estimates shows the drastic fall and rebound of economic activity during and after the lockdown of spring 2020. The combination of fat tails and heterogeneous dynamics allow our Bayesian DFM to pick up a sharp reduction and rapid recovery in activity. In fact, the model times the end of the 2020 recession at the end of April with a remarkable precision. It is also visible how the uncertainty around the estimates of the model increases quickly as the model revises upwards the volatility of the factor.

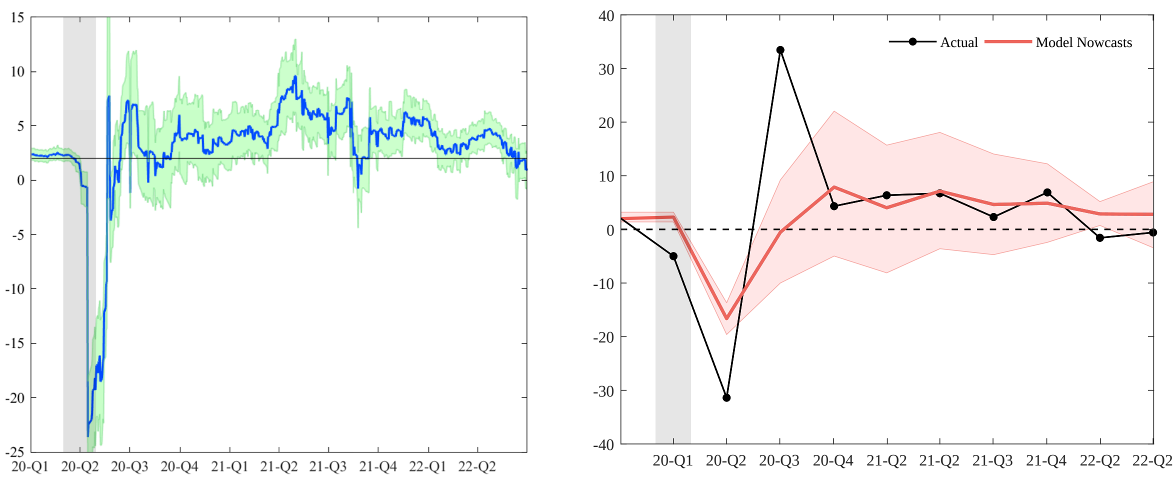
Panel (b) of Figure 10 shows how the model's daily assessment of the US economy translates into quarterly GDP nowcasts, by comparing predictions of the model in the middle of the quarter (in red) with the official GDP data subsequently published by the BEA (in black). The uncertainty around the model's estimates is expressed through the red shaded area. It is visible that despite the large swings that the model tracks at daily frequency in Panel (a), its quarterly

[30] It is worth noting that 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).

Figure 10: REAL-TIME ESTIMATES OF ACTIVITY DURING AND AFTER THE PANDEMIC

(a) Daily assessment of GDP growth

(b) Realized quarterly GDP growth and nowcasts



Notes. Panel (a) plots the daily estimate of US real GDP growth (blue line) with associated uncertainty (green shaded areas) from the beginning of 2020 onwards. Panel (b) shows, as the black line, the third release of real GDP growth in the United States as published by the BEA. The red line plots the nowcasts for real GDP growth based on information up to the mid-point of a given quarter (45 days). The shaded area indicate the 68% posterior credible intervals around the nowcasts.

nowcasts understate the magnitude of both the large reduction and large increase in the BEA releases for 2020:Q2 and 2020:Q3 GDP growth. In the recovery period, beginning with 2020:Q4, the model returns to producing reasonably accurate nowcasts, with uncertainty remaining elevated. The reason for the limited quarterly nowcasting ability during these two quarters appears to be that our model is allocating much of the rebound in GDP to the months of May and June 2020, i.e. *within* 2020:Q2. A closer look at Panel (a) makes clear that a swift recovery in activity is estimated to take place in June of 2020, and the daily estimate is back near the pre-pandemic average already by the end of the second quarter. This anticipation is related to the strong releases of the hard data series, such as retail sales as analyzed in Figure 9. The actual rebound in the national account data, however, occurred in July 2020 and the following months. So while our framework has a substantial edge over benchmark models over the 2000-2019 period, and captures well the qualitative developments of the pandemic, it is still limited in its ability to quantitatively match the magnitude and timing of the developments in real GDP during these two specific quarters. Nevertheless, its accuracy is re-established after 2020:Q3.

5.3 Large volatility, unstable comovement and alternative data

The COVID-19 episode is unique not only because of the massive increase in aggregate volatility, but also because of changes in the sectoral behavior of macroeconomic time series. The recession led to a large contraction in sectors that are usually not very cyclical, 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, decoupled series that usually comove. In our model, variables can deviate from their usual comovement 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. Which variables are important for the estimation of the factor thereby varies over time even if the loadings are fixed. In Appendix E we provide additional analysis in the form of variance decomposition, which illustrates the role of changes in comovements in more details.

Finally, several researchers have advocated using newly available high-frequency data to track the economy (e.g. Chetty et al., 2020). As these alternative data series display dynamics that are potentially different from standard macro data, modeling outliers and heterogeneous lead-lag dynamics might be especially critical in such applications.[31]

6 Conclusion

We propose a Bayesian DFM that incorporates 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. The explicit modeling of heterogeneous dynamics and fat tails substantially changes the way that information is processed by the model for the purposes of nowcasting, allowing for a more balanced view of which series are informative about economic activity compared to existing approaches. In a comprehensive evaluation exercise based on fully real-time, unrevised data between 2000 and 2019, the nowcasting performance is substantially

[31] An earlier version of this paper (Antolin-Diaz, Drechsel, and Petrella, 2021) incorporates newly available high-frequency data, such as credit card transactions, driving trips, or restaurant reservations. To address the issue that these data have a very short history, we propose a simple method to incorporate them into DFMs based on their close relation to existing traditional indicators. 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.

stronger than that of benchmark models, and comparable or better than that of professional human forecasters. The model also proves useful to track the macroeconomic developments during the challenging period of spring of 2020 and the subsequent recovery.

References

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

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

ANGELETOS, G.-M., F. COLLARD, AND H. DELLAS (2020): "Business-Cycle Anatomy," *American Economic Review*, 110, 3030–70.

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

——— (2021): "Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data," CEPR Discussion Papers 15926, C.E.P.R. Discussion Papers.

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., D. GIANNONE, AND L. REICHLIN (2011): "Nowcasting," in *Oxford Handbooks in Economics*, Oxford University Press, USA, vol. II, 193–224.

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.

BAŃBURA, M. AND G. RÜNSTLER (2011): "A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP," *International Journal of Forecasting*, 27, 333–346.

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

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

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

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

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 (2022a): "Nowcasting tail risk to economic activity at a weekly frequency," *Journal of Applied Econometrics*, 37, 843–866.

CARRIERO, A., T. E. CLARK, M. MARCELLINO, AND E. MERTENS (2022b): "Addressing COVID-19 outliers in BVARs with stochastic volatility," Working paper.

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," Ecb working papers 2453.

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.

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

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.

HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

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

——— (2004): "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185–212.

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.

JURADO, K., S. C. LUDVIGSON, AND S. NG (2015): "Measuring Uncertainty," *American Economic Review*, 105, 1177–1216.

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.

LENZA, M. AND G. E. PRIMICERI (2020): "How to Estimate a VAR after March 2020," NBER Working Papers 27771, National Bureau of Economic Research.

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.

NG, S. (2021): "Modeling Macroeconomic Variations After COVID-19," Working paper.

NORDHAUS, W. D. (1987): "Forecasting Efficiency: Concepts and Applications," *The Review of Economics and Statistics*, 69, 667–674.

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: The Role of Heterogeneous Dynamics and Fat Tails"

by Juan Antolin-Diaz, Thomas Drechsel and Ivan Petrella

Contents

A Details on model and algorithm

A.1 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 autoregressive 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, $\{a_t, f_t\}_{t=1}^T$, the parameters, $\boldsymbol{\theta}$, and the stochastic volatilities, $\{\sigma_{\varepsilon,t}, \sigma_{\eta_{i,t}}\}_{t=1}^T$ 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 f_t, are initialized by running the Kalman filter and smoother once conditional on the initialized parameters. Set $j = 1$.

1. Draw outlier and idiosyncratic components conditional on estimated common factors

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

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 and idiosyncratic components. This step is independent for each of the variables in the system as such it can be run in a univariate state-space and be parallelized. The univariate state spaces are all at monthly frequency, and in the case of quarterly variables the estimation of the states spaces also produces interpolated monthly values for the quarterly variables applying the approximation in Mariano and Murasawa (2003). The interpolated quarterly variables are used in later steps of the Gibbs sampler.

2. Draw the parameters of the outlier component

For each variable, $i = 1, \dots, N$, 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, with $\psi_{i,t}$ i.i.d. inverse gamma, or $v_i/\psi_{i,t} \sim \chi^2_{v_i}$, with $z_{i,t} \sim N(0, \sigma^2_{o,i})$ one has that $o_{i,t} = \sqrt{\psi_{i,t}} z_{i,t} \sim t_{v_i}(0, \sigma^2_{o,i})$. 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 $p(\psi^j_{i,t}|o^j_{i,t}, \sigma^j_{o,i}, v^{j-1}_i) \sim IG\left(\frac{v^{j-1}_i+1}{2}, \frac{2}{(o^j_{i,t}/\sigma^j_{o,i})^2+2}\right)$ The degree of freedom v_i are discrete with probability mass proportional to the product of t distribution ordinates $p(v^j_i|o^j_{i,t}, \sigma^j_{o,i}) = p(v)\prod_{t=1}^T \frac{v^{-1/2}\Gamma(v+1/2)}{\Gamma(1/2)\Gamma(v/2)}(v + (o^j_{i,t}/\sigma^j_{o,i})^2)^{-(v+1)/2}$, 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 the common factor and trend component conditional on model parameters and SVs

Obtain a draw $\{a^j_t, f^j_t\}^T_{t=1}$ from $p(\{a_t, f_t\}^T_{t=1}|\{o^j_{i,t}\}^T_{t=1}, \boldsymbol{\theta}^{j-1}, \{\sigma^{j-1}_{\varepsilon,t}, \sigma^{j-1}_{\eta_{i,t}}\}^T_{t=1}, \mathbf{y})$.

Having computed the outlier adjusted series (i.e. $\Delta y^{OA}_{i,t} = \Delta(y_{i,t} - o_{i,t})$) and interpolated the quarterly series in Step 1 of the algorithm, this step produces a draw of the entire state vector $\mathbf{x}_t$ (which includes the long-run growth component, a_t, and the common factor, f_t) of the state-space representation described in Section A.2, and using the precision filter as described in Section **??**. 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 $\omega^{2,j}_a$ from $p(\omega^2_a|\{a^j_t\}^T_{t=1})$.

Taking the sample $\{a^j_t\}^T_{t=1}$ drawn in the previous step as given, and posing an inverse-gamma prior $p(\omega^2_a) \sim IG(S_a, v_a)$ the conditional posterior of ω^2_a 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 $\mathbf{\Phi}^j$ from $p(\mathbf{\Phi}|\{f^j_t, \sigma^j_{\varepsilon,t}\}^T_{t=1})$.

Taking the sequences of the common factor $\{f^j_t\}^T_{t=1}$ and its stochastic volatility $\{\sigma^{j-1}_{\varepsilon,t}\}^T_{t=1}$ 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 as given the idiosyncratic components drawn in Step and the sequence for the stochastic volatility of the i^{th} component, $\{\sigma_{\eta_{i,t}}^{j-1}\}_{t=1}^T$, the standardized idiosyncratic component is obtaiend, which follows 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.

A.2 Construction of the state space system for Step 3 of the Gibbs sampler

For expositional clarity, in this section we abstract from the heterogeneous dynamics. 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 $\mathbf{y}_t$, which contains the (outlier adjusted and de-meaned) n_q interpolated quarterly and n_m monthly variables (i.e. $n = n_q + n_m$).[1] Therefore the time index t is always monthly, both for the quarterly and the monthly variables. The state space system is defined so that the system is written out in terms of the *quasi-differences* of the indicators, $\tilde{\mathbf{y}}_t$, defined as

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

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 $\mathbf{x}'_t = [a_t, a_{t-1}, a_{t-2}, f_t, f_{t-1}, f_{t-2}]$. We order GDP, GDI and consumption growth as the first three variables in $\tilde{\mathbf{y}}_\mathbf{t}$ and assume they share a common low frequency component. Therefore in Equation (2) we set $\mathbf{b} = (1\ 1\ b_c\ 0\ \ldots\ 0)'$. Setting $\lambda_1 = 1$ for identification, the matrices of parameters $\mathbf{H}$ and $\mathbf{F}$, are then constructed as shown below:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_a & \mathbf{H}_\lambda \end{bmatrix},$$

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

$$\mathbf{H}_a = \begin{bmatrix} 1 & -\rho_{1,1} & -\rho_{1,2} \\ 1 & -\rho_{2,1} & -\rho_{2,2} \\ b_c & -b_c\rho_{3,1} & -b_c\rho_{3,2} \\ & \mathbf{0}_{(n-3)\times 3} & \end{bmatrix}, \qquad \mathbf{H}_\lambda = \begin{bmatrix} 1 & -\rho_{1,1} & -\rho_{1,2} \\ \lambda_2 & -\lambda_2\rho_{2,1} & -\lambda_2\rho_{2,2} \\ \vdots & \vdots & \vdots \\ \lambda_n & -\lambda_n\rho_{n,1} & -\lambda_n\rho_{n,2} \end{bmatrix},$$

[1] The interpolation of the quarterly variables is performed in step 1 of the Gibbs sampler.

and

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_2 \end{bmatrix},$$

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

$$\mathbf{F}_1 = \left[\begin{array}{cc} 1 & \mathbf{0}_{1\times 2} \\ \hline \mathbf{I}_2 & \mathbf{0}_{2\times 1} \end{array}\right] \qquad \mathbf{F}_2 = \left[\begin{array}{ccc} \phi_1 & \phi_2 & 0 \\ \hline \multicolumn{2}{c}{\mathbf{I}_2} & \mathbf{0}_{2\times 1} \end{array}\right].$$

The innovations to the transition equation are denoted as

$$\mathbf{e}_t = \begin{bmatrix} v_{a_t} & \mathbf{0}_{2\times 1} & \epsilon_t & \mathbf{0}_{2\times 1} \end{bmatrix}',$$

with covariance matrix $\mathbf{Q}_t = diag(\omega_a^2, \mathbf{0}_{1\times 2}, \sigma_{\epsilon,t}^2, \mathbf{0}_{1\times 2})$. Whereas the covariance matrix of the measurement equation is defined as $\tilde{\mathbf{R}}_t = diag(\sigma_{\eta_{1,t}}^2, \ldots, \sigma_{\eta_{n,t}}^2)$.

B Additional in-sample results

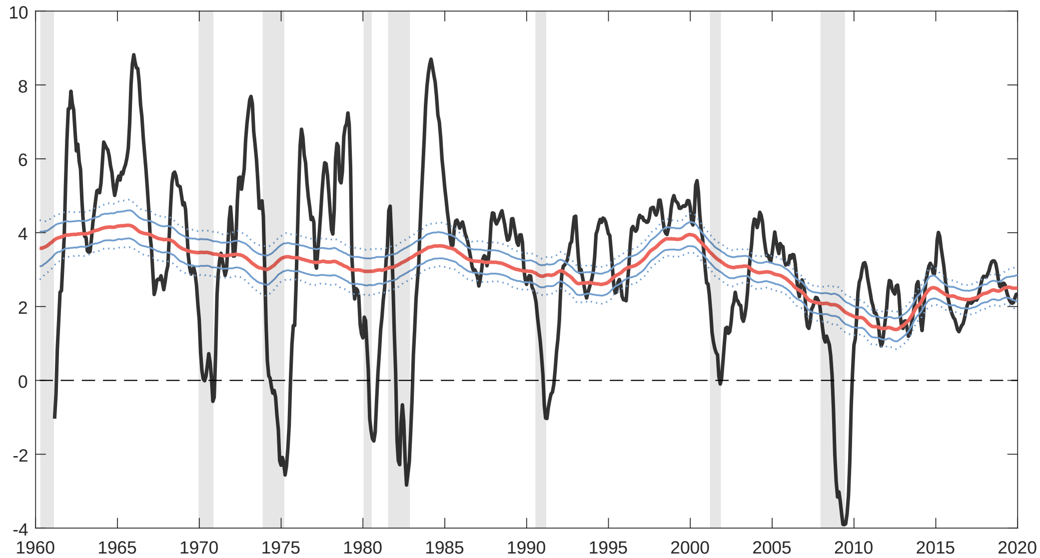
B.1 Secular trends: long-run growth and drifting volatilities

Figure B.1 displays the estimate of the time-varying long-run growth rate of GDP as well as the SV of the common factor. These estimates condition on the full sample and account for both filtering and parameter uncertainty. Panel (a) presents the posterior median 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).

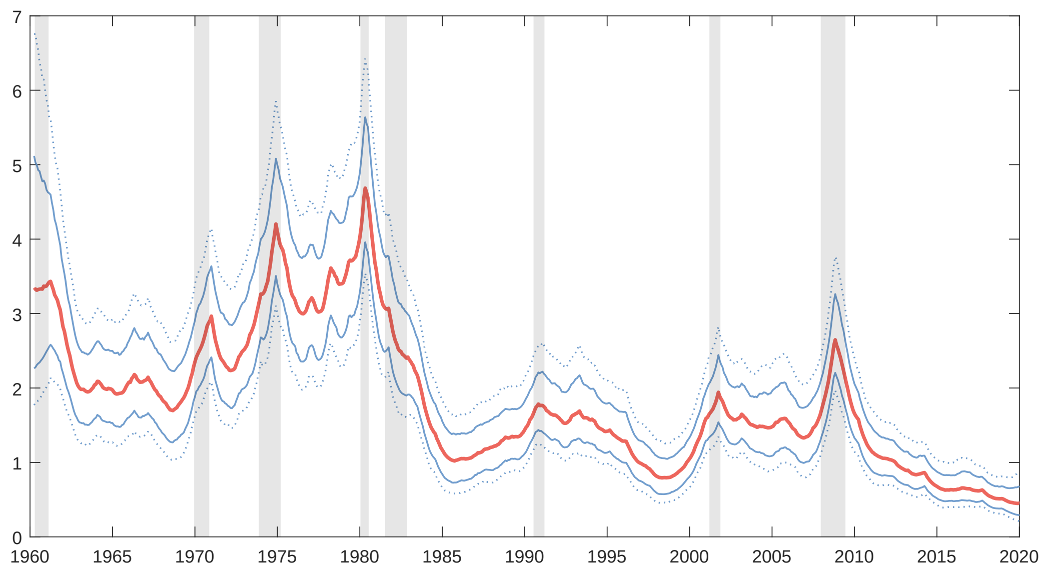
Panel (b) presents the posterior estimate of the SV of the common factor. Volatility declines over the sample, with the Great Moderation clearly visible and still in place. Just prior to the COVID-19 pandemic, output volatility reached a historically low level of less than 1% in annualized terms. The 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 ones.

Figure B.1: 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.

B.2 Uncertainty index, volatility and fat tails

In this section we compute an index of uncertainty following Jurado et al. (2015) using our model.[2] Figure B.2 reports the mean and HPD bands of the uncertainty index

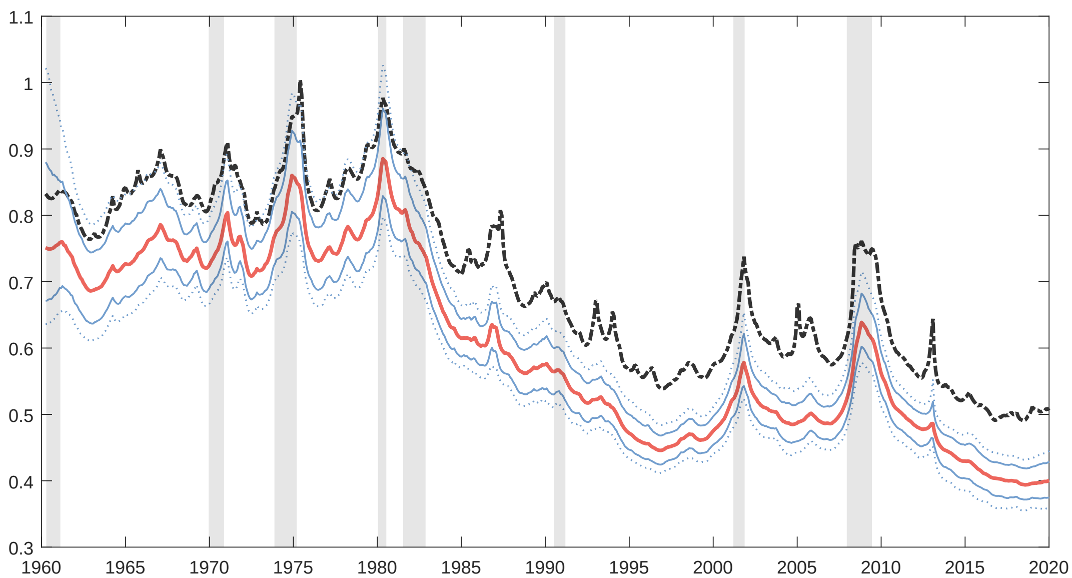
$$\mathcal{U}_t \equiv \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{\lambda_i^2 \sigma_{\varepsilon,t}^2 + \sigma_{u,i,t}^2}{\hat{s}_i^2}} \tag{15}$$

where $\hat{s}_i$ is a variable-specific scaling factor capturing differences in average standard 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 B.1 (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 explanation is our treatment of fat tails: the broken black line in Figure B.2 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 can be attributed to short-lived natural disasters such as hurricane Katrina in 2005. Our results suggest caution when interpreting economic uncertainty indices in the presence of fat-tailed innovations that may not wash out in the aggregate.[3]

[2] Online Appendix B presents the estimated SV of all idiosyncratic components of the model.

[3] 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).

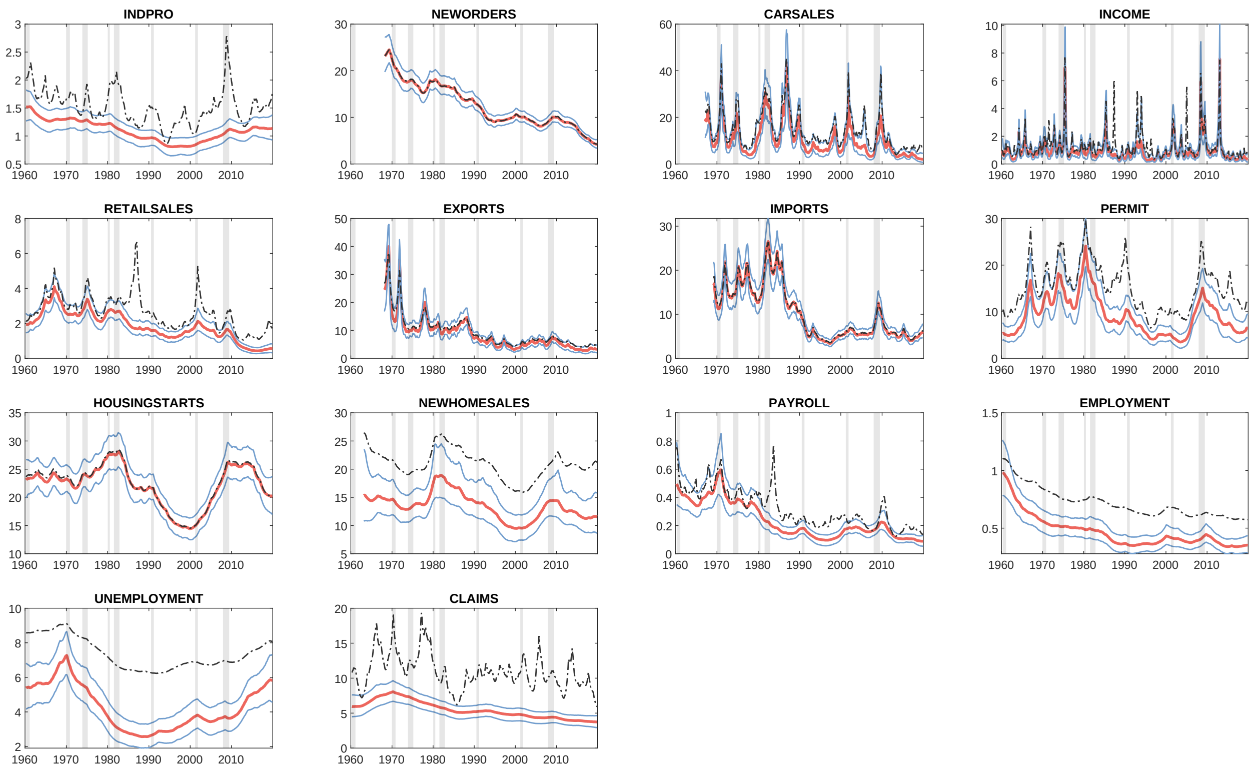
Figure B.2: UNCERTAINTY INDEX WITH AND WITHOUT OUTLIER TREATMENT



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 (15) 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.

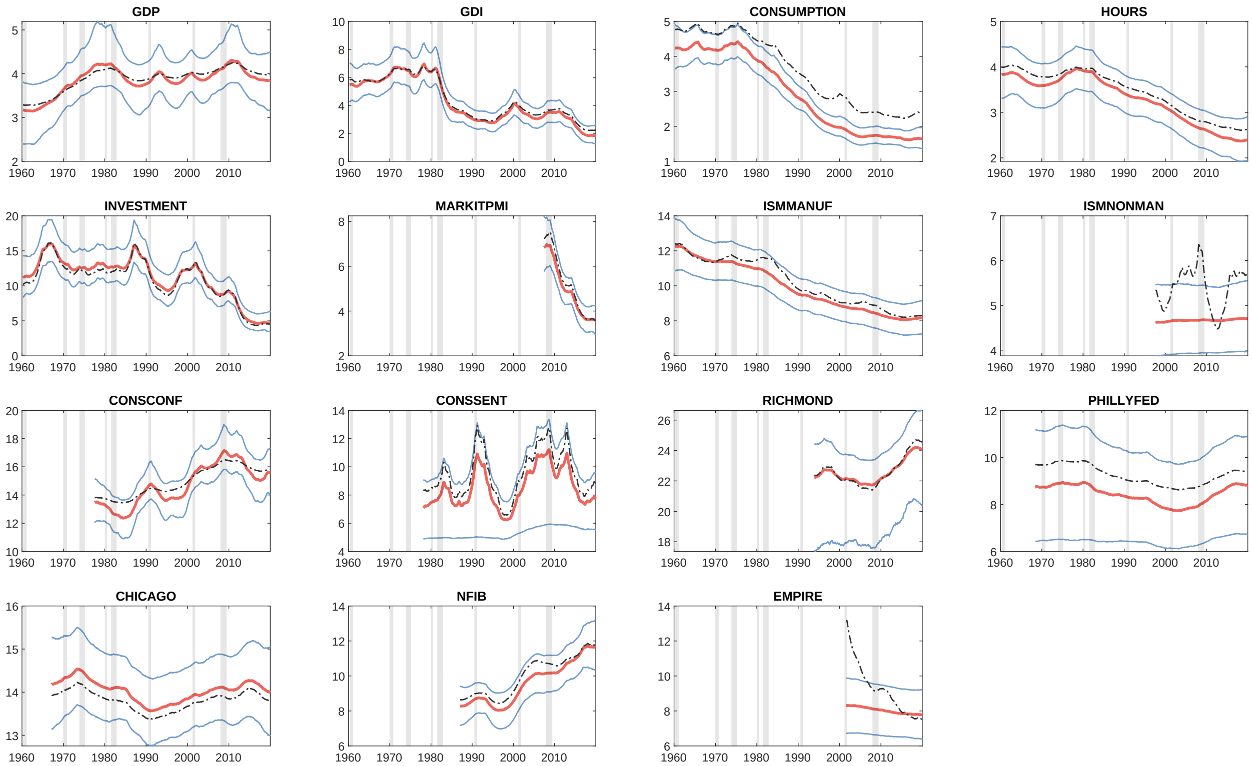
B.3 SV of idiosyncratic components

Figure B.3: POSTERIOR ESTIMATE OF SV OF MONTHLY HARD VARIABLES



Notes. Each panel presents the median (solid red), the 68% (solid 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 B.4 presents the quarterly variables as well as the monthly soft variables. Table 1 provides a full list of the individual data series with more details.

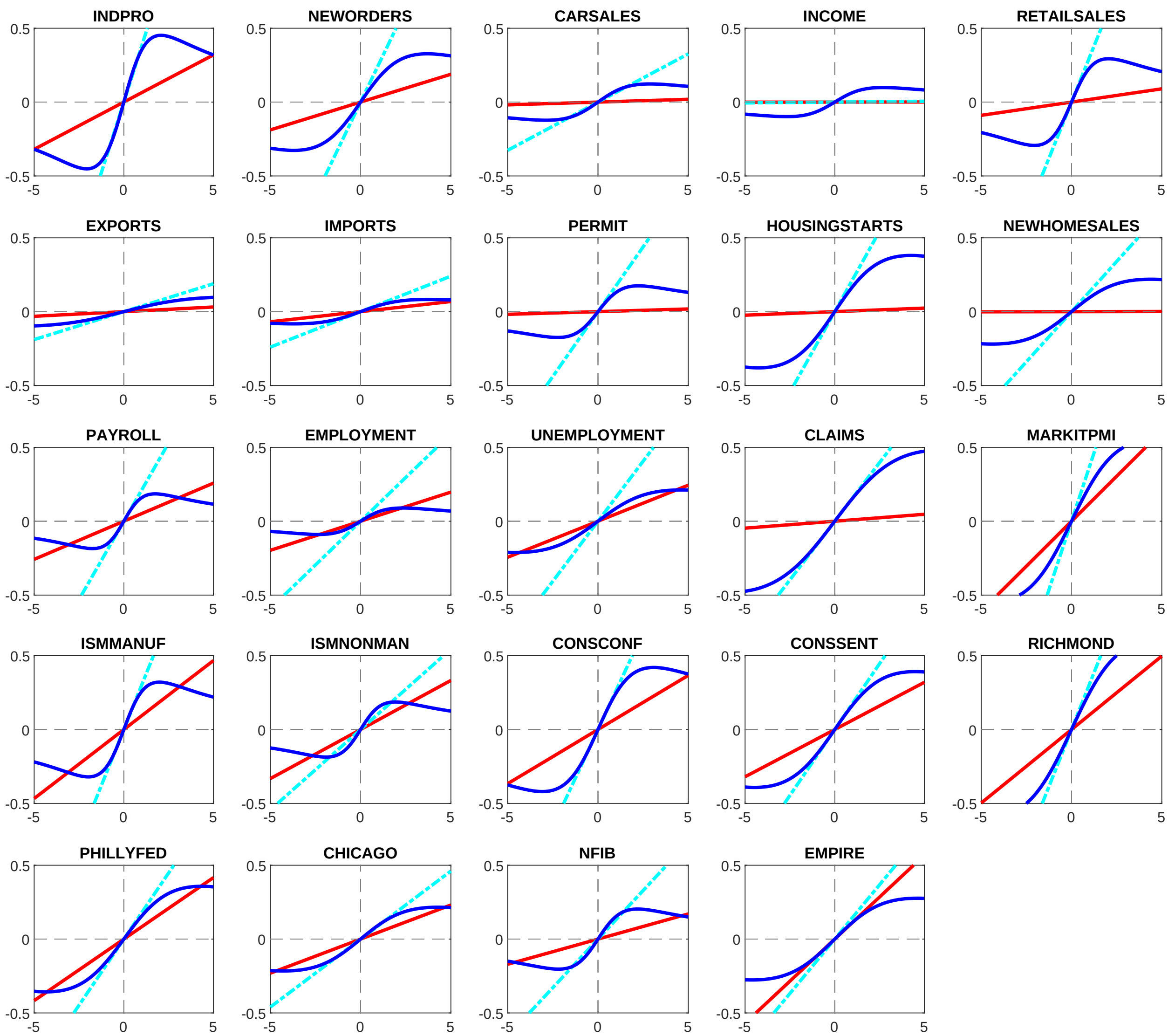
Figure B.4: POSTERIOR ESTIMATE OF SV OF QUARTERLY AND MONTHLY SOFT VARIABLES



Notes. Each panel presents the median (solid red), the 68% (solid 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 B.3 presents the monthly hard variables. Table 1 provides a full list of the individual data series with more details.

B.4 Influence function for all variables

Figure B.5: 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 (12) - (14) and explained in the main text, the model with fat tails allows these functions to be nonlinear and nonmonotonic.

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

C.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.

C.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.1, 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.[4] 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.

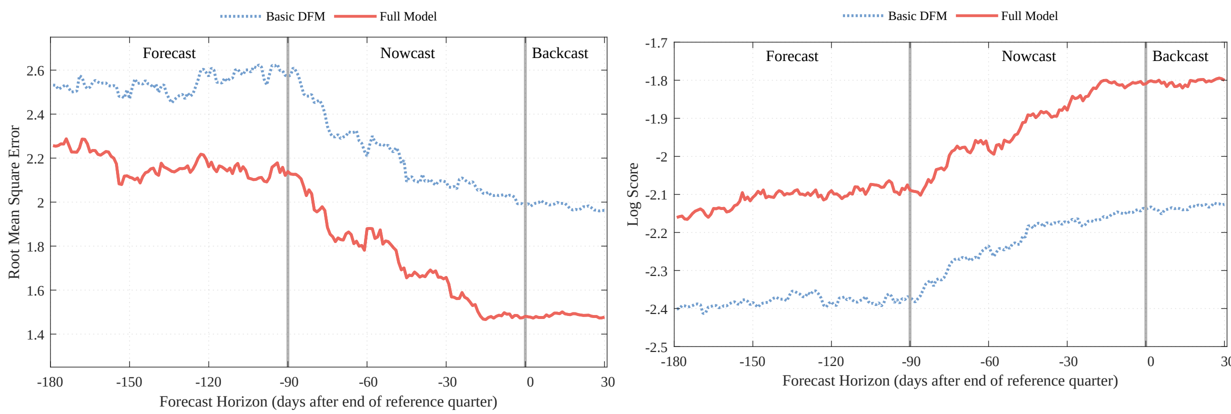
[4]This calculation is carried out using an Intel(R) Core(TM) i7-8700 CPU 3.20GHz with 32.0 GB of RAM.

D Additional forecast evaluation results

D.1 Forecast evaluation as the data flow arrives

Figure D.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.

Figure D.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 2 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 D.3.

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.

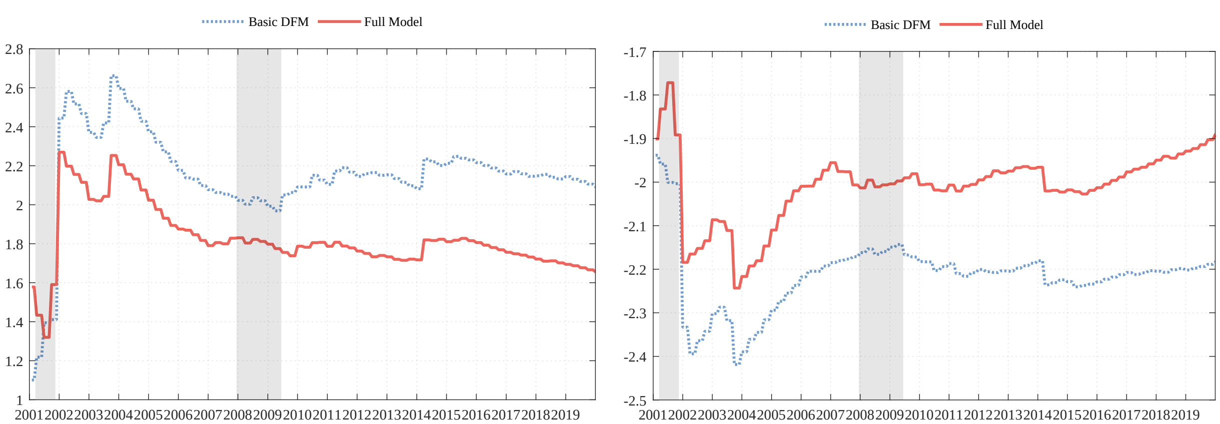
Panel (b) turns to density forecasts, so evaluates the accuracy of the entire predictive distribution, instead of focusing exclusively on its center. They are used to predict unusual developments or tail risks, such as the probability of a recession or a strong recovery given current information. Our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

D.2 Forecast evaluation through time

Figure D.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.

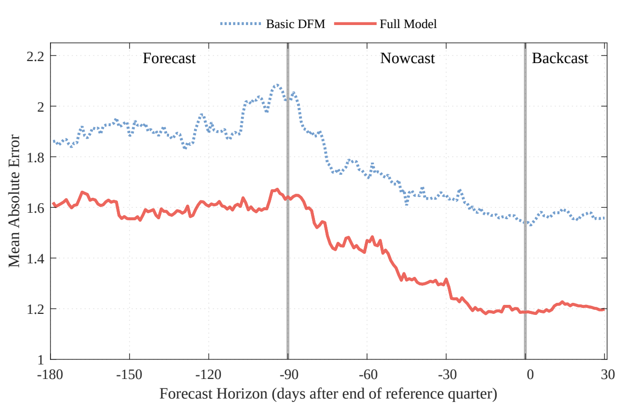
The results in Figure D.1 document the average performance over the period 2000-2019. It is useful to examine whether the out-performance of the full model is stable over time or due to a few special periods. In Figure D.2 we present, for a fixed forecast horizon, the relevant loss function computed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

D.3 Alternative metrics for forecast evaluation

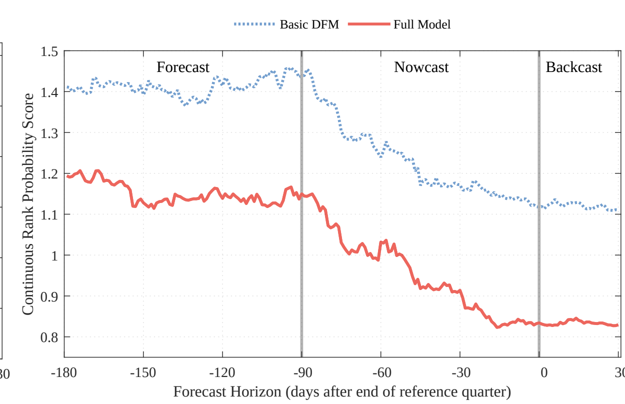
This Appendix examines the robustness of our formal forecast evaluation for alternative evaluation metrics. Figures D.3 and D.4 present the results shown in Figures D.1 and D.2, using alternative evaluation metrics. Figure D.3 focuses on the evaluation across horizons (as the data flow arrives) and Figure D.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 D.3: FORECAST EVALUATION AS THE DATA FLOW ARRIVES (ALTERNATIVE METRICS)

(a) Point forecast accuracy: MAE



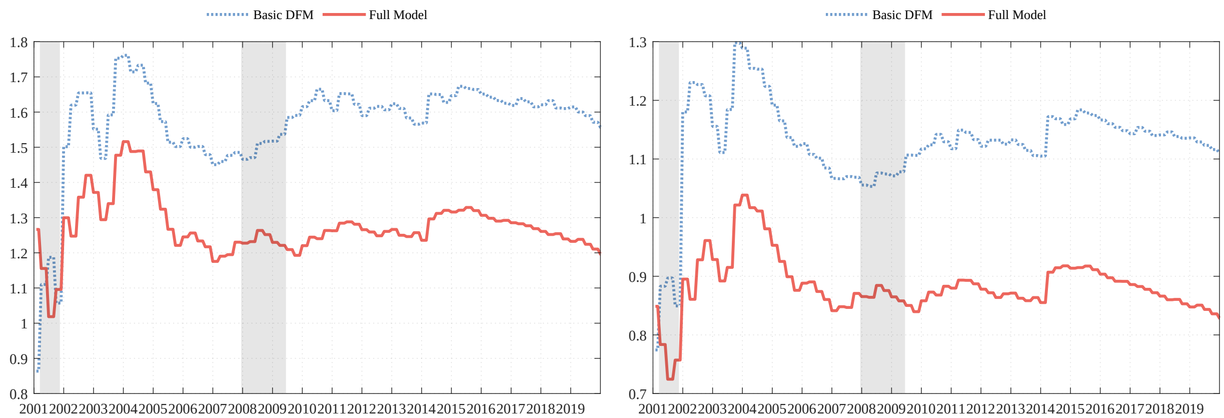
(b) Density forecast accuracy: CRPS



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 D.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.

D.4 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 D.5 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 2 in the main text, but provide a graphical representation over a continuous evolution of the forecast horizon. Furthermore, relative to Table 2 the figure breaks down the performance of our model into the contribution of the individual components: trends and SV, heterogeneous dynamics and fat tails. 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.

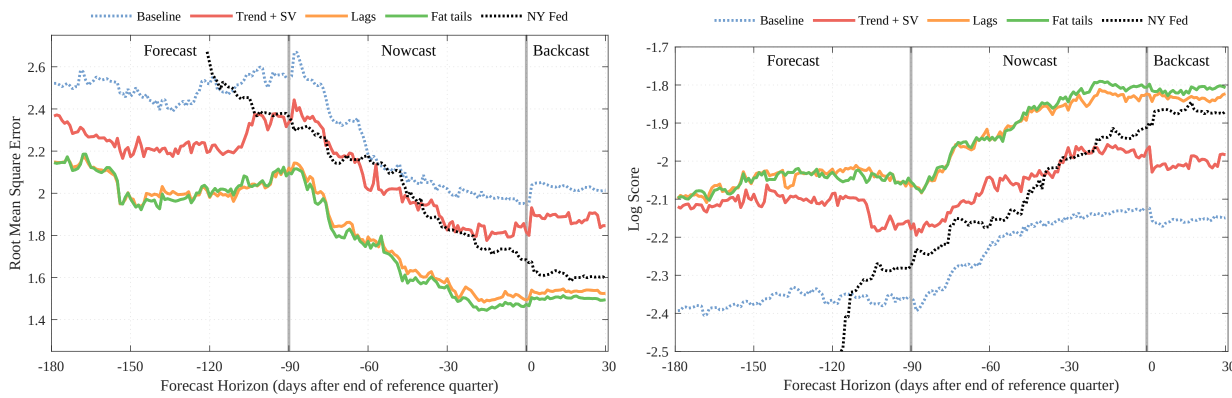
The results shown in Figure D.5 document the average performance over the period 2000-2019. Figure D.6 instead presents, for a fixed forecast horizon, the relevant loss function computed recursively *over time*. In other words, this figures extends the analysis in Appendix D.2 to a finer breakdown of model versions and includes the NY Fed's nowcasting model.

D.5 Details on using the Survey of Professional Forecasters

In Section 4.4 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

Figure D.5: 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).

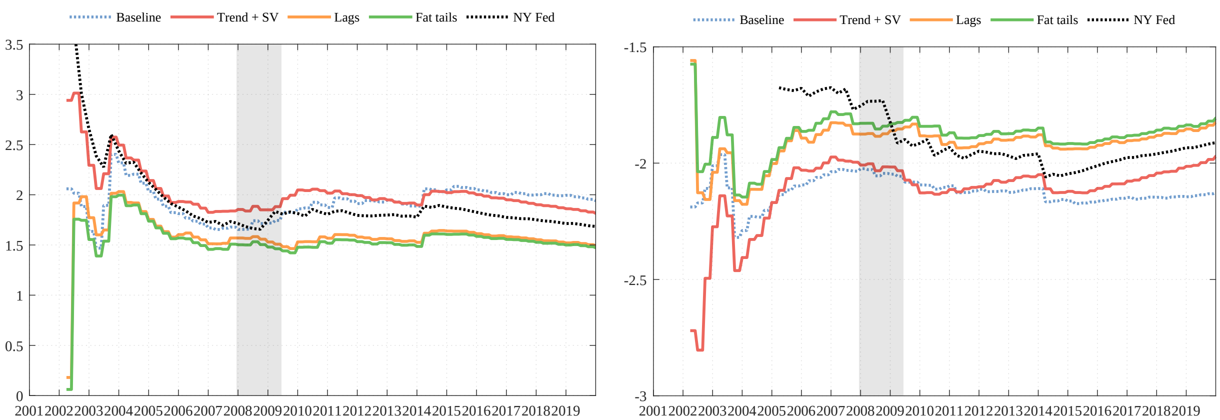
background on the construction of the data required for that comparison, as well some additional analysis of the relative performance of our model for multiple period forecasts.

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 D.7 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

Figure D.6: 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).

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.

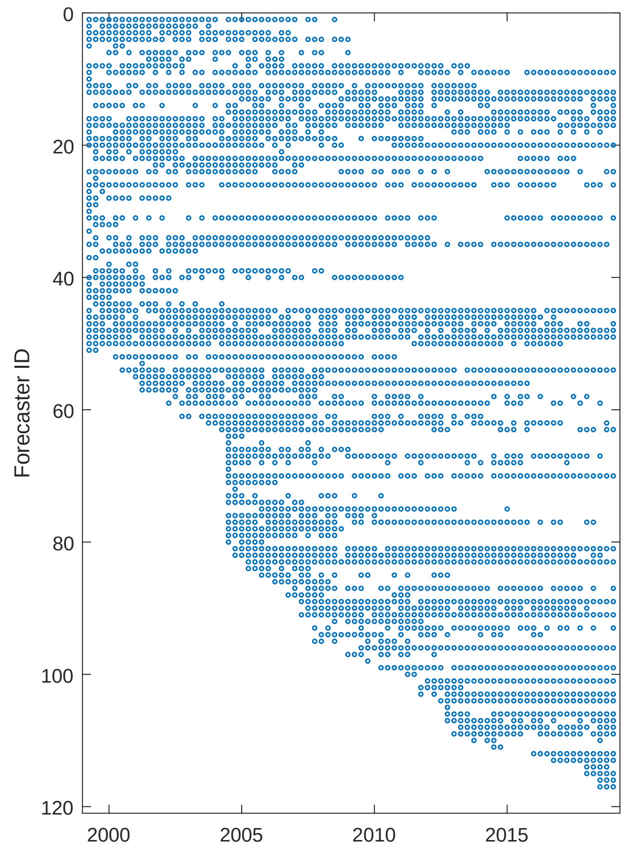
D.6 Real-time assessment of activity, uncertainty, tail risks

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.[5]

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 D.8 plots three real-time measures of risk. Panel (a) displays a real-time measure of uncertainty, defined

[5]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.

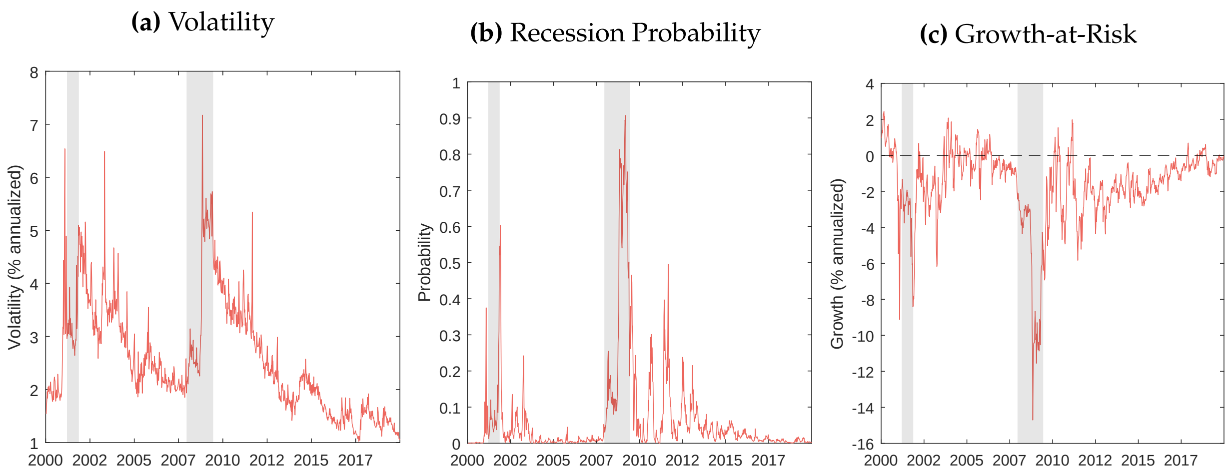
Figure D.7: AVAILABILITY OF INDIVIDUAL SPF FORECASTS



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.

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 B.1, but computed in real time. Therefore, it can be interpreted as a measure of business cycle uncertainty as perceived at each point in time by an observer with access to our model. Panel (b) shows at the probability of recession, defined as the model-implied probability that the current and next quarter GDP growth will be negative. Finally, panel (c) presents a measure of GDP growth at risk in the spirit of Adrian et al. (2019), measuring the mean below the 5% distribution of the GDP distribution for the current quarter. All these measures are useful characterizations of the risks around economic activity that go beyond the information contained in central estimates. The good performance in density forecast exhibited by the full model gives us confidence that they can be relied upon in practice.

Figure D.8: REAL-TIME RISK ASSESSMENT MEASURES



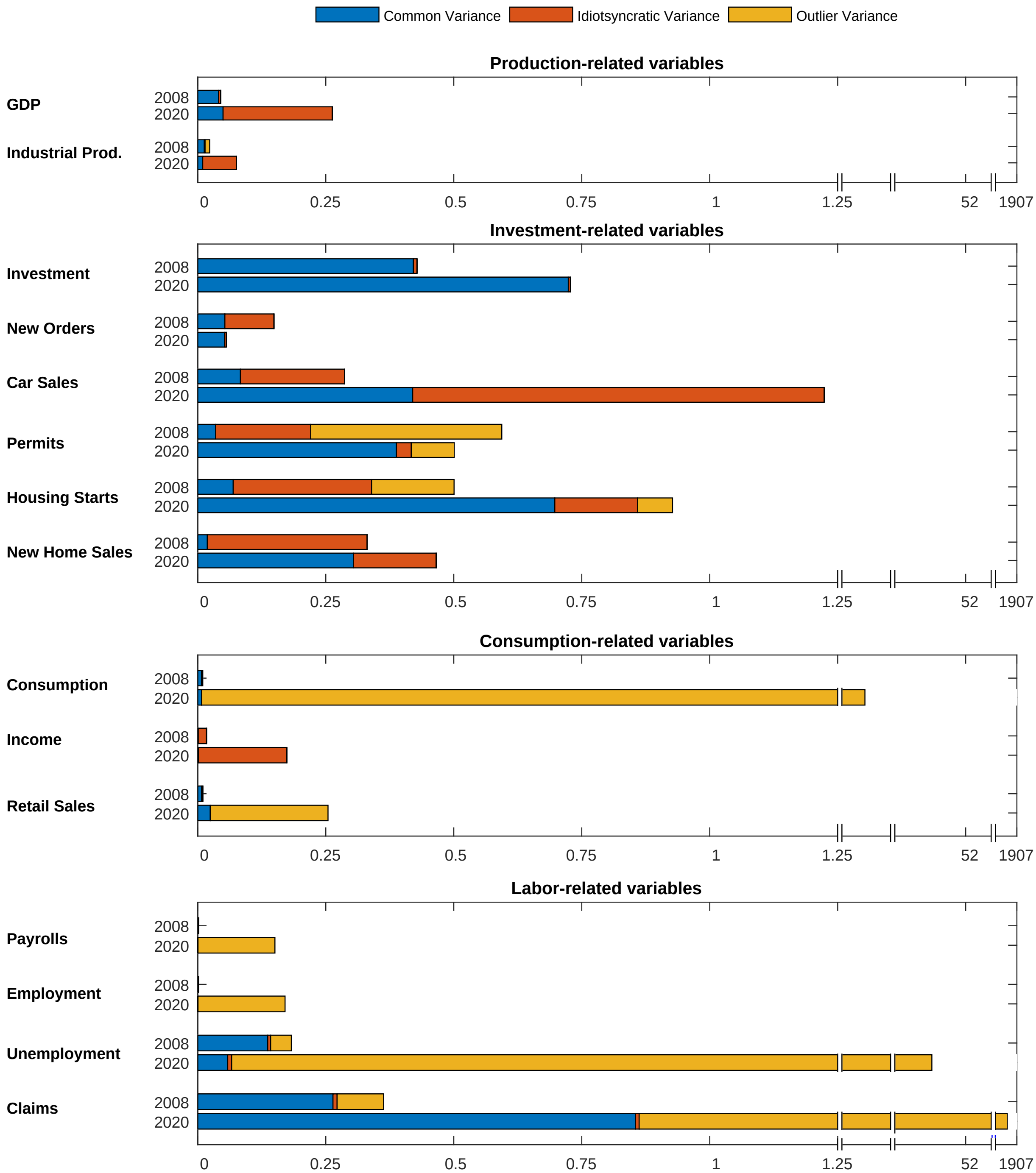
Notes. Panel (a) displays a real-time measure of uncertainty, defined as the difference between the 16th and the 84th percentiles of the daily activity estimate. Panel (b) plots the probability of recession, defined as the model-implied probability that the current and next quarter GDP growth will be 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.

E Analysis of the comovement of macro variables in 2020

This appendix highlights how the comovement between macroeconomic variables changed dramatically in the recession of 2020, and how our modeling innovations, in combination, can deal with this important feature of the episode. To this end, we analyze a variance decomposition of the variables included in out model for particular episodes in the sample. 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.[6] Figure E.1 reports such a variance decomposition for selected variables during 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. 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.

[6] 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 E.1: 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.

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.

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.

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.

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