International Journal of Forecasting 39 (2023) 1460–1476

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Nowcasting GDP with a pool of factor models and a fast estimation algorithm☆

Sercan Eraslan [a], Maximilian Schröder [b,c,*]

[a] Deutsche Bundesbank, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany
[b] BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway
[c] Norges Bank, Bankplassen 2, P.O. Box 1179 Sentrum, 0107 Oslo, Norway

ARTICLE INFO

Keywords:
Dynamic factor model
Forecasting
Mixed-frequency
Model averaging
Time-varying parameter
Stochastic volatility

ABSTRACT

We propose a novel mixed-frequency dynamic factor model with time-varying parameters and stochastic volatility for macroeconomic nowcasting and develop a fast estimation algorithm. This enables us to generate forecast densities based on a large space of factor models. We apply our framework to nowcast US GDP growth in real time. Our results reveal that stochastic volatility seems to improve the accuracy of point forecasts the most, compared to the constant-parameter factor model. These gains are most prominent during unstable periods such as the Covid-19 pandemic. Finally, we highlight indicators driving the US GDP growth forecasts and associated downside risks in real time.

1. Introduction

The Covid-19 pandemic was accompanied by unusually large fluctuations in key macroeconomic indicators and thus poses new modeling challenges. At the same time, policymakers such as central banks require a robust assessment of the past, current, and future state of economic activity to conduct informed, forward-looking, and responsible monetary policy. Naturally, the recent forecasting literature has been interested in exploring modeling approaches that account for these extreme outliers and help robustify the performance of forecasting models in abnormal times (see e.g. Carriero et al., 2021, 2022; Lenza & Primiceri, 2020; Schorfheide & Song, 2021).

The macroeconomic forecasting literature considers several additional modeling challenges. First, different economic indicators are often released at asynchronous dates, different frequencies, and with publication lags. Against this backdrop, Mariano and Murasawa (2003) develop a mixed-frequency dynamic factor model (MF-DFM) that can handle such data characteristics. By formulating a factor model, the authors also mitigate the risk of parameter proliferation, which might lead to higher computational demand and imprecise predictions when datasets grow large. Other examples include Bańbura et al. (2013, 2010), Bańbura and Modugno (2014), and Giannone et al. (2008).

Second, a changing economic environment can spark changes in the economic transition mechanism, the comovement of variables, and the nature and heteroskedasticity of exogenous shocks. Del Negro and Otrok (2008) and more recently Marcellino et al. (2016) specify factor models with time-varying parameters (TVP) and stochastic volatility (SV) to reduce their vulnerability to structural breaks. Examples in the forecasting literature include

☆ We thank the editor and two anonymous referees as well as Dimitris Korobilis and Leif A. Thorsrud for their valuable comments and suggestions. This paper was circulated under the title "Nowcasting GDP with a large factor model space" in earlier versions.

This paper represents the authors' personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

* Corresponding author.
E-mail addresses: sercan.eraslan@bundesbank.de (S. Eraslan), maximilian.schroder@bi.no (M. Schröder).

https://doi.org/10.1016/j.ijforecast.2022.07.009

Clark (2011), Clark and Ravazzolo (2015), D'Agostino et al. (2013), and Marcellino et al. (2016). The authors find that models with stochastic volatility (SV) provide more accurate point and density forecasts than, e.g., constant-parameter models. Similarly, the Covid-19 literature proposes SV models to account for extreme data realizations (see e.g. Carriero et al., 2021). Compared to homoskedastic VARs, the authors find that forecasts generated from SV VARs are less sensitive to outliers during the pandemic.

Third, the same factors that drive parameter change might also alter the forecast performance of economic predictors and forecasting models over time. In periods of large, economy-wide disruptions—such as those observed during the pandemic—this might pose an additional modeling challenge. Evidence for the existence of these "pockets of predictability"[1] for a range of economic time series is provided by Byrne et al. (2018), Dangl and Halling (2012), Farmer et al. (2018), Koop and Korobilis (2012), and Rossi (2013). A potential solution is to use performance-based model averaging schemes that allow forecast combinations to change dynamically. In this context, Raftery et al. (2010) propose dynamic model averaging (DMA), which is successfully applied by Koop and Korobilis (2011) and Onorante and Raftery (2016) in the context of macroeconomic forecasting.

While the above-mentioned studies address some of the challenges individually, a unified modeling approach seems to be missing. Similarly, recent studies investigate whether individual modeling choices lead to forecast performance gains during the pandemic, but do not encompass TVPs, SV, and model averaging simultaneously. The question of whether any combination of these three techniques promises incremental forecast improvements, especially during the pandemic, thus also remains unanswered. We seek to fill these gaps in the literature.

We propose a novel time-varying-parameter mixed-frequency dynamic factor model with stochastic volatility (TVP-MF-DFM-SV). Our model is fully integrated into a dynamic model averaging framework for macroeconomic nowcasting. In doing so, we contribute to the nowcasting literature in various ways. First, we build a TVP-MF-DFM-SV that can efficiently deal with the properties of real-time data flow and time variation in the parameters, as well as volatility. Regarding these characteristics, our model is closely related to that proposed by Thorsrud (2020). Instead of applying standard Bayesian techniques, however, we follow a different estimation strategy, which leads to our second contribution. We develop a fast, dual one-step Kalman filter algorithm which only requires a single iteration. Therefore, we extend the algorithm proposed by Koop and Korobilis (2014) to account for mixed-frequency data. This algorithm enables us to estimate a large space of factor models in a reasonable amount of time. Third, we estimate our TVP-MF-DFM-SV in a unified dynamic model averaging framework, where nowcasts are based on many different model specifications. This framework also sheds light on the time-varying importance of the economic indicators and hence on the drivers of forecasts and associated downside risks. This might be of paramount interest to policymakers.

In our empirical analysis, we apply various specifications of the proposed TVP-MF-DFM-SV-DMA framework to forecast quarterly GDP growth in the US. Our recursive out-of-sample real-time forecasting exercise reveals valuable insights into the marginal benefits of time-varying parameters, stochastic volatility, and dynamic model averaging. Our results show that SV models provide marginal gains compared to the constant-parameter benchmark in stable times. In a more volatile environment, SV models provide sizeable gains, while TVP-SV models, and to a lesser extent TVP models, also provide performance improvements for some forecast horizons. When paired with dynamic model averaging, further forecast performance gains can be observed. In addition, compared to fully Bayesian benchmarks, the results demonstrate the competitiveness of our proposed forecasting model and the robustness of our approximate estimation algorithm. When highlighting which predictors drive the downside risks associated with our GDP growth forecasts, we find this exercise to produce intuitive results. For instance, during the Covid-19 pandemic interest rates, CPI, consumer sentiment, new orders, and retail sales emerge as dominant drivers. In contrast, during the Great Recession stock prices and housing indicators seem to be most strongly associated with downside risks. Finally, we extract recession probabilities from our model space and find that the model's predictions largely match the business-cycle turning points announced by the NBER. Altogether, our approach offers a model that yields precise point forecasts in expansions and recessions along with the variables that emerge as important drivers of the US GDP growth forecast distribution across the model space.

The remainder of this paper is organized as follows. The next section introduces the econometric framework and the estimation algorithm. Section 3 provides an overview of the forecasting exercise and presents its results. The final section concludes.

2. Econometric methodology

In this paper, we utilize a novel mixed-frequency dynamic factor model with time-varying-parameters and stochastic volatility in a dynamic model averaging framework for nowcasting US GDP growth. Therefore, we first briefly describe an MF-DFM, which is the point of departure of our proposed factor model. Then, we introduce the extensions of the baseline model, namely time-varying parameters (TVPs) and stochastic volatility (SV). Subsequently, we present the dynamic model averaging (DMA) framework. Finally, we take the reader through the proposed estimation algorithm.

2.1. MF-DFM

Throughout the following, upper-case letters indicate matrices and lower-case letters indicate vectors. Further, we adopt the convention that a superscript M (Q) indicates model components that are related to monthly

[1] This terminology was introduced by Farmer et al. (2018).

(quarterly) variables. Let x_t denote the n-dimensional zero-mean vector of stationary variables. We now assume that, at monthly frequency, the economy linearly depends on k latent factors, f_t, that are common to all variables contained in x_t. This can be written as

$$x_t = \Lambda_t \cdot f_t + u_t, \quad u_t \sim N(0, V_t). \tag{1}$$

The so-called common component $(\Lambda_t \cdot f_t)$ captures the variability in the dependent variables that is due to the common factors, whereas the idiosyncratic zero-mean Gaussian disturbances, u_t, account for the remaining variability. We assume the latter to satisfy $E(f_t u_t) = 0$ and to be cross-sectionally and serially uncorrelated.[2] To allow for a possible interaction between the factors, we define a dynamic process that is given by a pth-order VAR:

$$f_t = B_{t,1} f_{t-1} + \cdots + B_{t,p} f_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim N(0, Q_t) \tag{2}$$

where ε_t denotes serially uncorrelated zero-mean Gaussian disturbances. Instead of extracting factors and using them to augment univariate forecasting regressions, we thus model the variables jointly in a multivariate system. This should improve the identification of the variables' comovements and thus benefit the forecasting results (see Koop & Korobilis, 2014).

As pointed out above, economic indicators are usually observed at different frequencies. While GDP, the key indicator of interest in our forecasting exercise, is available at quarterly frequency, other macroeconomic indicators such as industrial production and the unemployment rate are observed on a monthly basis. This results in a dataset, x_t, consisting of mixed-frequency indicators. To this end, we specify the factor model given in Eqs. (1) and (2) at monthly frequency and define the relationship that links the quarterly variables to their latent high-frequency counterparts following Bańbura et al. (2013, 2010), and Mariano and Murasawa (2003). This relationship critically hinges on whether the variables of interest are stock or flow variables and on how they have been transformed before entering the model. In case of a quarterly flow variable such as GDP, one has the accounting identity

$$Y_t^Q = Y_t^M + Y_{t-1}^M + Y_{t-2}^M, \tag{3}$$

where Y_t^M denotes the unobserved monthly counterpart of Y_t^Q during the respective quarter. After transforming the observed Y_t^Q by applying log-differences one can define the partially observed monthly series:

$$y_t^Q = \begin{cases} log(Y_t^Q) - log(Y_{t-3}^Q), & t = 3, 6, 9... \\ unobserved, & otherwise, \end{cases} \tag{4}$$

where y_t^Q is observed every third month and unobserved during the first and second months of every quarter. We can now apply the approximations in Mariano and Murasawa (2003) and combine Eqs. (3) and (4), leading to the final aggregation scheme for the log-differenced flow variable y_t^Q:

$$\begin{aligned} y_t^Q &\approx log(Y_t^Q) - log(Y_{t-3}^Q) \\ &\approx \sum_{i=0}^{2} log\left(Y_{t-i}^M\right) - \sum_{i=0}^{2} log\left(Y_{t-i-3}^M\right) \\ &= y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \end{aligned} \tag{5}$$

for $t = (3, 6, 9...)$, where $y_t = \Delta log(Y_t^M)$. These approximations keep the constraints on the observational relationship linear and allow us to cast the model into a linear state-space representation (see Bańbura et al., 2013).

2.2. TVP-MF-DFM-SV

Having described the aggregation scheme and the baseline MF-DFM, we introduce time variation in the parameters (TVPs), as well as in the variance–covariance matrices (SV). Considering the restrictions which the aggregation scheme imposes on the structure of the factor model, we cast the system into state-space form[3]:

$$x_t = H_t s_t + u_t, \qquad u_t \sim N(0, V_t) \tag{6a}$$

$$s_t = A_t s_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, Q_t) \tag{6b}$$

$$\lambda_t = \lambda_{t-1} + v_t, \qquad v_t \sim N(0, W_t) \tag{6c}$$

$$\beta_t = \beta_{t-1} + \eta_t, \qquad \eta_t \sim N(0, R_t) \tag{6d}$$

with

$$x_t = \begin{bmatrix} y_t^M & y_t^Q \end{bmatrix}',$$

$$H_t = \begin{bmatrix} \Lambda_t^M & 0 & 0 & 0 & 0 & 0_{(m \times p-5)} \\ \Lambda^Q & 2\Lambda^Q & 3\Lambda^Q & 2\Lambda^Q & \Lambda^Q & 0_{(q \times p-5)} \end{bmatrix},$$

$$s_t = \begin{bmatrix} f_t & f_{t-1} & \dots & f_{t-4} & \dots & f_{t-p+1} \end{bmatrix}'$$

with $\Lambda_{(q \times k)}^Q$, $\Lambda_{(m \times k)}^M$, and

$$A_t = \begin{bmatrix} B_{1,t} & \dots & B_{p-1,t} & B_{p,t} \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix}, \qquad u_t = \begin{bmatrix} u_{1,t} \\ \vdots \\ u_{N,t} \end{bmatrix},$$

$$\varepsilon_t = \begin{bmatrix} e_t & 0 & \dots \end{bmatrix}'.$$

The vector of stationary variables, x_t, consists of y_t^M (y_t^Q) denoting the m (q) variables that are originally observed at monthly (quarterly) frequency. The corresponding loading matrices, Λ_t^M and Λ_t^Q, are contained in H_t. The state vector is denoted by s_t and restricted to have at least five elements as implied by the aggregation scheme. Its evolution follows a dynamic process which is governed by companion form matrix A_t. Finally, the zero-mean Gaussian disturbances of the measurement and transition

[2] Bańbura and Modugno (2014) find that explicitly accounting for serial correlation does not lead to consistent improvements in GDP forecasts. Thus, we do not expect this simplification to influence our results noticeably.

[3] We briefly introduce the Kalman filter and smoother equations in Online Appendix B.1.

equations are given by u_t and ε_t, where V_t is diagonal[4] and Q_t is singular with a non-singular block in the upper-left corner.

In case of the above model, Eqs. (6c) and (6d) govern the time variation in the factor loadings Λ_t and the parameter matrices of the dynamic factor process $B_{t,i}$, with $\lambda_t = vec(\Lambda_t)$ and $\beta_t = (vec(B_{t,1})', \ldots, vec(B_{t,p})')'$. We assume that both vectors evolve as multivariate random walks. Moreover, v_t and η_t are serially uncorrelated and feature the time-varying covariance matrices W_t and R_t. All disturbance vectors are further assumed to evolve independently.

In this setup, the innovations in Eq. (6a) are independent across the variables in x_t conditional on knowing s_t. Together with the assumption that the loadings are uncorrelated across variables, they can thus be sampled equation-by-equation. This allows for the construction of separate Kalman filter estimates of the factor loadings of the monthly (λ_t^M) and quarterly indicators (λ^Q). Note that the loading matrix of the quarterly variables, Λ^Q, is static. This is due to the fact that time variation in the loadings and the aggregation scheme described by Eq. (5) are in conflict with each other (see Thorsrud, 2020). Therefore, the estimation of λ^Q is not straightforward. Relying on, e.g., Bayesian regression requires knowledge of the variances, V_t^Q. To estimate V_t^Q as outlined below, however, we require an estimate of λ^Q at every point in time. Thus, a recursive method to estimate λ^Q is needed. Moreover, the sequence of quarterly variables is only partially observed and thus implies a dependent variable with many missing values. To work around this problem, conditional on the factors, we rotate the state space in Eq. (6a) and factor out Λ^Q. The relationship is now expressed as the product of the quarterly loadings and a moving average of the factors, such that

$$\begin{aligned} y_t^Q &= \Lambda_t^Q \cdot \sum_{i=0}^{4} \omega_i f_{t-i} + u_t, \quad u_t \sim N(0, V_t) \\ \omega_i &= \begin{cases} i+1, & for\ i = 0, 1, 2 \\ 6-i-1, & for\ i = 3, 4. \end{cases} \end{aligned} \tag{7}$$

We then manipulate Eq. (6c) to read $\lambda_t^Q = \lambda_{t-1}^Q$ and estimate the static quarterly loadings recursively by means of a Kalman filter. As quarterly variables are only observed in the last month of every quarter, they leave missing observations in other months. Accordingly, an update of the static parameter given the new information set occurs whenever y_t^Q is observed. When it is unobserved, we simply do not update λ^Q. In our estimation framework, which is described below, the implementation is particularly simple by setting $W_t = 0$ for the quarterly variables. Finally, as our smoothed estimate of λ^Q, we accept the most recent update for all time periods.

[4] This prevents the existence of a continuum of observationally equivalent models which are defined by arbitrary factor dependencies (see Nakajima & West, 2013). Since the algorithm employed in a later section relies on principal component estimates as starting values for the factors, the model is identified up to a sign rotation without having to impose further restrictions on the loading matrix (see Koop & Korobilis, 2014).

Similarly, we estimate the VAR coefficients β_t in the transition equation by means of a Kalman filter and smoother along similar lines to Koop and Korobilis (2014). In doing so, we only accept non-explosive draws. The Kalman filter recursions, as given in Online Appendix B.1, then proceed as usual.

We now shift our focus to the time variation in the variance–covariance matrices V_t, Q_t, W_t, and R_t that are introduced in Eqs. (6a)–(6d). In order to recursively estimate these matrices, we make use of simulation-free variance discounting methods. Starting with V_t and Q_t, we follow Koop and Korobilis (2014) and use exponentially weighted moving average (EWMA) estimators, given by

$$V_t = \kappa_1 V_{t-1} + (1-\kappa_1) \cdot diag(u_t u_t') \tag{8a}$$

$$Q_t(k) = \kappa_2 Q_{t-1}(k) + (1-\kappa_2) e_t e_t' \tag{8b}$$

where $u_{i,t}$ and $e_{i,t}$ are backed out according to Eqs. (6a) and (6b). With slight abuse of notation, $Q_t(k)$ denotes the kth leading principal submatrix of Q_t, and $diag(u_t u_t')$ denotes the diagonal matrix of the variance–covariance matrix of the residuals u_t. We use the diagonal matrix in Eq. (8a) to enforce the standard identifying constraint spelled out above and use the kth leading principal submatrix in Eq. (8b) to restrict the EWMA to the nonsingular block of Q_t. The degree of time variation in V_t and Q_t is governed by the two decay parameters, κ_1 and κ_2, respectively. As pointed out above, the mixed-frequency structure of our data introduces periodically missing values in the quarterly variables. Thus, the residuals necessary to compute the EWMA for V_t^Q are only available at the end of each quarter. We hence follow West and Harrison (1997), who provide suggestions on the treatment of missing values in EWMA and only update when actual data are available. Throughout the quarter, the EWMA does not decay and remains at its value.[5] This update lag, however, results in slower time variation of V_t^Q than V_t^M, even given the same κ_1. To compensate for this effect, in our empirical exercise, we allow V_t^Q and V_t^M to change at different rates and introduce the decay parameters κ_1^Q and κ_1^M. Finally, we follow Koop and Korobilis (2014) and produce smoothed estimates of V_t^Q, V_t^M, and Q_t.

Compared to standard approaches, this specification of stochastic volatility is an unfortunate necessity that one has to accept in order for the fast one-step algorithm to work. Nonetheless, this compromise seems justifiable, because EWMA estimators produce minimum mean-squared-error forecasts (see Muth, 1960) that are equivalent to those produced by simple state-space or ARIMA models (see Durbin & Koopman, 2012). In addition, as Koop and Korobilis (2014) point out, EWMA estimators provide an accurate approximation of integrated GARCH models and are thus in line with the features of the macroeconomic VAR literature that usually works with integrated stochastic volatility models (see e.g. Primiceri, 2005). Despite relying on this rather simple algorithm, this allows us to stay relatively close to the

[5] Since the Kalman filters treat missing information similarly, this also has the appeal of maintaining a consistent approach to the treatment of missing data.

standard methods. Nonetheless, in light of time periods that are characterized by sudden spikes in volatility, such as the Covid-19 pandemic or the global financial crisis, this specification might seem somewhat restrictive. In our empirical exercise, we thus update the forgetting factors with every GDP release. In unstable times such as the Covid-19 pandemic, this might result in a lower forgetting factor, which translates to past observations receiving a lower weight. Compared to more stable periods, this allows volatility to spike more abruptly. At the same time, if the forgetting factor is too flexible during stable periods, this introduces noise, and the out-of-sample forecasting performance might suffer. In this case, updating the forgetting factors allows us to adopt a smoother transition in the model's parameters. To lend support to this idea, Figure D.1 in the Online Appendix displays the in-sample volatility of the residual of GDP extracted from the model space (introduced below). Both gradual changes and considerable spikes in volatility (e.g. during the global financial crisis or the Covid-19 pandemic) can indeed be observed. This provides some empirical evidence in support of this modeling choice.

For W_t and R_t, we use the forgetting-factor methods described in Raftery et al. (2010) and Koop and Korobilis (2012). Thus, we estimate these matrices directly from the respective state covariance matrix estimates provided by the Kalman filter. From standard Kalman filter inference, we know that λ_t and β_t in Eqs. (6c) and (6d) are given by

$$\lambda_t|Data_{1:t-1} \sim N(\lambda_{t|t-1}, \Sigma^{\lambda}_{t|t-1}) \tag{9a}$$

$$\beta_t|Data_{1:t-1} \sim N(\beta_{t|t-1}, \Sigma^{\beta}_{t|t-1}) \tag{9b}$$

where by Eqs. (6c) and (6d)

$$\Sigma^{\lambda}_{t|t-1} = \Sigma^{\lambda}_{t-1|t-1} + W_t \tag{10a}$$

$$\Sigma^{\beta}_{t|t-1} = \Sigma^{\beta}_{t-1|t-1} + R_t. \tag{10b}$$

Following Raftery et al. (2010) and Koop and Korobilis (2014), one can now define $W_t = (\kappa_3^{-1} - 1)\Sigma^{\lambda}_{t-1|t-1}$ and $R_t = (\kappa_4^{-1} - 1)\Sigma^{\beta}_{t-1|t-1}$ to replace Eqs. (10a) and (10b) by

$$\Sigma^{\lambda}_{t|t-1} = \kappa_3^{-1}\Sigma^{\lambda}_{t-1|t-1} \tag{11a}$$

$$\Sigma^{\beta}_{t|t-1} = \kappa_4^{-1}\Sigma^{\beta}_{t-1|t-1} \tag{11b}$$

which introduces the forgetting factors κ_3 and κ_4, respectively. Hence, the model is still a properly defined state space, and the Kalman filters and smoothers proceed in standard fashion. The interpretation of the decay parameters and forgetting factors is generally the same. Lower values put lower weight on past observations and thus allow for faster parameter change. A value of one implies constant parameters. In our empirical application, we use this feature to estimate nested baseline MF-DFM model specifications. This allows us to assess the relative gains of more complex models with time variation in the parameters and the variance–covariance matrices within a consistent framework.

2.3. Dynamic model averaging

While a specific model or some economic indicators might predict GDP particularly well during phases of stable growth, they might lose predictive power e.g., in times of economic crises, and vice versa. This seems particularly intuitive in periods of large, economy-wide disruptions such as those observed during the Covid-19 pandemic. More generally, the same factors that drive parameter change, such as policy decisions, the demise of industrial sectors, or technological advancement, might alter the forecast performance of economic predictors naturally over time. Ample evidence for the existence of these "pockets of predictability" for various economic time series is provided by Byrne et al. (2018), Dangl and Halling (2012), Farmer et al. (2018), Koop and Korobilis (2012) and Rossi (2013). Consequently, no single model with a constant set of predictors can be expected to beat all its competitors continuously, and variable choices should be updated on a regular basis (see Banerjee et al., 2005). In a factor model context, Boivin and Ng (2006) and Bańbura and Modugno (2014) provide evidence that changes in data heterogeneity also imply changes in the overall factor structure. As a result, while more sparse and homogenous data might require fewer factors, more dense or heterogeneous data might require more factors. At the same time, this suggests that the number of factors and with it the factor structure evolve over time.

Against this backdrop, we estimate a multitude of factor model specifications of different sizes and apply forecast combination methods to generate point and cross-sectional density forecasts. Specifically, we define

$$x_t^{(j)} = H_t^{(j)} s_t^{(j)} + u_t^{(j)}, \qquad u_t^{(j)} \sim N(0, V_t^{(j)}) \tag{12a}$$

$$s_t^{(j)} = A_t^{(j)} s_{t-1}^{(j)} + \varepsilon_t^{(j)}, \qquad \varepsilon_t^{(j)} \sim N(0, Q_t^{(j)}) \tag{12b}$$

$$\lambda_t^{(j)} = \lambda_{t-1}^{(j)} + v_t^{(j)}, \qquad v_t^{(j)} \sim N(0, W_t^{(j)}) \tag{12c}$$

$$\beta_t^{(j)} = \beta_{t-1}^{(j)} + \eta_t^{(j)}, \qquad \eta_t^{(j)} \sim N(0, R_t^{(j)}) \tag{12d}$$

where superscript $j \in [1, \ldots, M]$ denotes a single model specification. Each of these models is uniquely identified by the combination of variables and the number of factors. In particular, we explore every possible variable combination for up to $\bar{k}$ factors. This amounts to $\bar{k} \cdot 2^{n-\bar{k}}$ forecasting models in total.

In our suggested approach, we apply dynamic model averaging (DMA) proposed by Raftery et al. (2010) to produce forecast combinations. Briefly, DMA is a recursive updating method that produces an estimate of how applicable each model in the forecasting model space is through time. This sets it apart from methods proposed more recently—by, e.g., Elliott et al. (2013) and Kim and Swanson (2014)—that largely ignore model instability. The fact that all M models can be estimated independently allows for a very simple and computationally fast approximation that involves updating the models' probabilities individually at each point in time (see Raftery et al., 2010).

In the general DMA framework, $\pi_{t|t-1,j}$ denotes the probability that model j applies at time t, conditional on the information set being valid through $t-1$. Since it is not reasonable to believe that this probability is independent of the forecast horizon h, we extend the general framework and define $\pi_{t|t-1,j,h}$. Given that some variables might be more informative for longer forecast horizons,

while others are more suitable short-term predictors, this allows for assigning forecast-horizon-specific weights to the models. Lastly, it remains to be defined how the individual model probabilities evolve through time. Raftery et al. (2010) propose the following approximations:

$$\pi_{t|t-1,j,h} = \frac{\pi^{\gamma}_{t-1|t-1,j,h}}{\sum_{k=1}^{J} \pi^{\gamma}_{t-1|t-1,k,h}} \tag{13a}$$

$$\pi_{t|t,j,h} = \frac{\pi_{t|t-1,j,h} AFE_j^{-1}(I_t|I_{1:t-h})}{\sum_{k=1}^{J} \pi_{t|t-1,k,h} AFE_k^{-1}(I_t|I_{1:t-h})} \tag{13b}$$

where Eqs. (13a) and (13b) are the prediction and the update step, respectively, and $\gamma \in [0, 1]$ is another forgetting factor that controls the rate of time variation in the individual model probabilities and thus model change. A higher (lower) γ implies slower (faster) model change. For $\gamma = 1$, the model probabilities are analogous to those under Bayesian model averaging (BMA). In a more general modeling approach, one would have to estimate the transition matrix that governs the system directly. If the model space grows large, however, the computational demand implied by standard methods quickly becomes prohibitive. The appeal of the proposed approximation is that this can be avoided by specifying the transition matrix indirectly through a forgetting factor. For more details, we refer the interested reader to Online Appendix B.2 and Raftery et al. (2010).

$AFE_j(I_t|I_{1:t-h})$ denotes the mean absolute deviation (AFE) of model j at time t given past information and serves as a measure of fit. Note that the general framework utilizes the predictive likelihood instead (see Koop & Korobilis, 2014). Since we are interested in point forecasts, however, we deviate to keep the framework consistent. The model probabilities thus evolve conditionally on the past point forecast performance of the individual models at time t and for forecast horizon h. The combined DMA forecasts then arise as the model probability weighted average of the M individual forecasts. Finally, we construct forecast densities from the cross-section of the models' point forecasts. Our cross-sectional forecast densities are thus distinct from classical simulated forecast densities.

Besides its practical advantages, DMA offers distinct theoretical features that are appealing in our context. By assigning weights to individual models based on their past forecast performance and thus indirectly to the models' parameters, DMA implies shrinkage. Especially in larger models, this serves to counteract overparameterization (see Koop & Korobilis, 2011). In our setup, where the model space contains variable as well as factor combinations, shrinkage through DMA also helps to address dynamics in the factor structure that are driven by time variation or data heterogeneity. If a more homogeneous variable combination forecasts better with fewer factors, DMA assigns a higher weight compared to the same variable combination with more factors. Similarly, if this relationship reverses over time, this change should again be reflected in the DMA weights. Given that our DMA weights are also horizon-specific, this even allows for the case in which the required number of factors for a given variable combination might depend on the length of the forecast horizon. If fewer factors lead to superior forecast performance for longer forecast horizons, models with fewer factors should in turn also be assigned larger weights at longer horizons. Overall, DMA thus leaves it to the data to decide which or how much of each model—and thus the implied factor structure—is desirable at each point in time.

2.4. Estimation algorithm

Models such as the TVP-DFM of Del Negro and Otrok (2008) or the MF-DFM-SV proposed by Marcellino et al. (2016) are usually estimated with Bayesian methods involving Markov chain Monte Carlo (MCMC) algorithms, such as Gibbs samplers. The obvious drawback of this procedure is high computational demand. Adding to this, the Kalman filters and multivariate stochastic volatility models that are usually used to assess the conditional distributions of the time-varying parameters and covariance matrices have to be re-estimated at each iteration of the algorithm. The estimation of even a single TVP-MF-DFM-SV by means of MCMC schemes is thus computationally costly. When faced with a recursive forecasting exercise on an expanding window of data or many different model specifications, as is the case with the model averaging techniques we employ in our empirical exercise, the computational demand again multiplies and quickly becomes prohibitive. In a real-world scenario, where a policymaking institution cannot afford to wait several days or even weeks for a forecast to be produced, this renders MCMC schemes inapplicable in our setup.

Instead, for the parameter estimation of our model, we develop a fast, dual one-step Kalman filter algorithm that only requires one single iteration, building on the one proposed by Koop and Korobilis (2014). The general idea is to circumvent the need for recursive sampling by conditioning on principal component estimates for the factors during the estimation of the model parameters, and by replacing the multivariate stochastic volatility models that are usually used for V_t, Q_t, W_t, and R_t by the variance discounting methods that are described above (see e.g. Aguilar & West, 1998). These changes break up the recursiveness of the Gibbs algorithm and thus allow for simulation-free estimations of the time-varying parameters and factors using Kalman filters. The treatment of missing observations is straightforward in this setting, where we follow Mariano and Murasawa (2003). Essentially, at every point in time, we induce the Kalman filter to skip missing observations such that only the observed variables affect the update of the state vector and its variance. Each vintage for each model j, our estimation algorithm evolves as follows:

(1) Initialization

   (a) Initialize the forgetting factors via grid search
   (b) Standardize the data
   (c) Estimate the preliminary factors f_t^{PC} by the EM algorithm.[6]

[6] Considering the mixed-frequency structure of our data, we estimate the preliminary factors from x_t using the expectation-maximization (EM) algorithm proposed by Stock and Watson (2002b).

(2) Parameter estimation

   (a) Estimate V_t, Q_t, W_t, and R_t using variance discounting methods
   (b) Estimate λ_t^M and β_t conditional on V_t, Q_t, W_t, R_t, and f_t^{PC}
   (c) Estimate λ^Q conditional on V_t, Q_t, W_t, R_t, and f_t^{PC}

(3) Factor estimation

   (a) Estimate f_t conditional on the model parameters

(4) Forecasting

   (a) Generate forecasts based on estimated factors and model parameters
   (b) Calculate model weights for dynamic model averaging

Every vintage, we initialize the forgetting factors via a grid search and estimate preliminary principal component factors by means of an EM algorithm. Conditioning on principal component estimates that disregard the time series structure of the system might be viewed as a potential shortcoming of our procedure. The efficiency loss induced by this choice might, however, be moderated in light of the results in Bates et al. (2013) and Stock and Watson (2002a). The authors demonstrate the robustness of principal components analysis (PCA) under various temporal instabilities, such as time-varying loadings and serial correlation of the factors, as well as residuals that are serially correlated, cross-correlated, and conditionally heteroskedastic. The non-static forgetting factors allow for an automatic adjustment of the speed of parameter change over time, which allows for more sudden adjustments in e.g., unstable periods.

Subsequently, we estimate the variance–covariance matrices using variance discounting methods. The model parameters are estimated based on the initial factor estimates and forgetting factors using Kalman filters and smoothers. Then, we estimate the factors conditional on the model parameters and variance–covariance matrices from the previous step. Finally, we generate forecasts for each model j and apply DMA to combine the model-specific predictions. The large computational gains of our procedure naturally come at a cost. Compared to its computationally demanding fully Bayesian counterparts, parameter estimation uncertainty, especially regarding the forgetting factors, is largely ignored. Given that our primary focus is on forecasting, we accept this caveat in order to develop a model that accounts for structural breaks, time-variation in the volatilities, mixed frequencies, and large information sets. To explore the extent to which this choice and the value of the forgetting factors in particular impact the forecast performance, we run a robustness check in Section 3.

3. Forecasting exercise

3.1. Data

We use 19 monthly and quarterly indicators for nowcasting GDP in the US. In addition to the target variable of our forecasting exercise—namely, quarterly US GDP growth—we include 18 monthly indicators selected from the eight groups in McCracken and Ng (2016) and provide timely information about the US economy. The data transformations follow the recommendations provided by the authors. Table 1 provides an overview.

Our dataset may appear somewhat small at the first glance, although factor models are able to deal with many more predictors. In this regard, however, Alvarez et al. (2012) show for key US macroeconomic variables that a small-scale factor model based on a set of representative indicators is able to outperform factor models containing a large set instead. Similarly, Bańbura and Modugno (2014) construct a small dataset, including industrial production, orders, retail sales, the unemployment rate, economic sentiment, GDP, the prices of industrial raw materials, stock prices, and the PMI, which is then extended with more disaggregated information to create increasingly large datasets. In their work, the authors demonstrate that a small factor model containing 14 variables can outperform a large model with 101 indicators for nowcasting GDP in the euro area. In another recent study, Duarte and Süssmuth (2018) start with 258 monthly indicators of Spanish GDP, which they reduce to a set of 14 final predictors prior to estimation. The final list of indicators includes subaggregates of industrial production and manufacturing, consumer and industry surveys, a financial index, and an employment indicator. While the absence of well-known predictors such as the PMI might be viewed as a caveat, our selected variables nonetheless show a large degree of overlap with the aforementioned studies and span a similarly wide range of indicator groups.

From our dataset, we generate a large factor model space. In order to ensure the continuity of GDP forecasts and the minimum number of indicators necessary for factor extraction, we include a tiny set of indicators—GDP, industrial production, manufacturing and trade sales, personal income ex. transfer, and total nonfarm payrolls—in every model specification.[7] Then, we extend the fixed set of variables with every possible combination of the remaining 14 indicators in our dataset and specify models with one, two, and three factors. This leads to $3 \cdot 2^{14} = 49{,}152$ possible model specifications.

3.2. Initial conditions and starting values

Since our algorithm is not fully Bayesian, we set initial conditions rather than actual priors. Nonetheless, we follow Thorsrud (2020) and choose relatively informative initial values for the variances and uninformative initial values for the remaining parameters.

$$f_0 \sim N(0, 10) \tag{14a}$$

$$\lambda_0^M \sim N(0, 1 \cdot I_n) \tag{14b}$$

$$\lambda_0^Q \sim N(0, 1 \cdot I_n) \tag{14c}$$

[7] Such indicators are commonly used to generate coincident indexes for the US business cycle, see e.g. Mariano and Murasawa (2003) and Stock and Watson (1991), among others.

Table 1
Dataset.

Indicator	Group	Freq.	Tra.
Real personal income ex. transfer	Output and income	M	log diff.
Industrial production	Output and income	M	log diff.
Capacity utilization	Output and income	M	1st diff.
Unemployment rate	Labor market	M	1st diff.
Initial claims	Labor market	M	log diff.
Total nonfarm payrolls	Labor market	M	log diff.
Housing starts	Housing	M	logs
Housing permits	Housing	M	logs
Real manu. and trade sales	Cons., orders, and inventories	M	log diff.
Retail and food sales	Cons., orders, and inventories	M	log diff.
New orders for capital goods	Cons., orders, and inventories	M	log diff.
Consumer sentiment	Cons., orders, and inventories	M	1st diff.
Federal funds rate	Interest and exchange rates	M	1st diff.
3-month Treasury bill	Interest and exchange rates	M	1st diff.
10-year Treasury rate	Interest and exchange rates	M	1st diff.
Oil price	Prices	M	2nd log diff.
CPI	Prices	M	2nd log diff.
S&P 500	Stock market	M	log diff.
GDP	Output	Q	log diff.

Notes: This tables gives an overview of indicators used in our study. The first and second columns display the indicator and its group. The column "Freq." denotes their frequency, which can be either monthly (M) or quarterly (Q). Data transformations presented in the last column are defined as 1st, 2nd, and log differences, or as logs only. While the monthly indicators are obtained from the FRED-MD introduced by McCracken and Ng (2016), quarterly GDP is retrieved from the ALFRED.

$$\beta_0 \sim N(\mu_{Min}, \sigma_{Min}) \tag{14d}$$

$$V_0 \equiv 0.1 \cdot I_n \tag{14e}$$

$$Q_0 \equiv 0.1 \cdot I_k \tag{14f}$$

$$\pi_{0|0,j,h} = \frac{1}{J}, \quad \text{for } j = 1, \ldots, J \tag{14g}$$

μ_{Min} and σ_{Min} indicate a Minnesota-style initial parameter distribution. Usually, the idea behind a Minnesota prior is to express beliefs about the structure of the VAR for the factor state equation, where more distant lags are penalized more strongly. We assume that the factor VAR follows a relatively persistent AR(1) with a coefficient of 0.9 as an initial condition for the respective Kalman Filter. The variance is given by $\sigma_{Min} = 0.1/r^2$ for the coefficient on lag r, which corresponds to the choice of Koop and Korobilis (2014).

What remains is the specification of the forgetting factors. To the best of our knowledge, such a model has not been estimated in this framework before, so we use a relatively simple grid search to guide our choice. First, we restrict the forgetting factors such that κ_1^M and κ_2, as well as κ_3 and κ_4, change at the same rate. This seems reasonable, because both pairs are defined at the same frequency and govern the parameter change of similar components. The decay parameter of the residual variances for quarterly variables κ_1^Q, however, is allowed to change at its own rate, due to reasons discussed above. Thus, depending on the considered model class, our grid has up to three dimensions. In general, the decay parameters can take values in the following intervals:

$$\kappa_1^M, \kappa_1^Q, \kappa_2 \in \{0.60, \quad 0.65, \quad \ldots, \quad 0.95, \quad 0.99\}$$
$$\kappa_3, \kappa_4 \in \{0.90, \quad 0.91, \quad \ldots, \quad 0.99\}$$

As an example, the TVP-MF-DFM-SV requires the full three-dimensional grid for all of the forgetting factors to be determined. By contrast, the MF-DFM specifications do not require a grid search, because the forgetting factors are set to $\kappa_1^M = \kappa_1^Q = \kappa_2 = \kappa_3 = \kappa_4 = 1$. For the TVP-MF-DFM model specification, we set $\kappa_1^M = \kappa_1^Q = \kappa_2 = 1$ and obtain κ_3 and κ_4 via grid search, whereas setting $\kappa_3 = \kappa_4 = 1$ yields the MF-DFM-SV specification. For each model class, we evaluate the grid for one, two, and three factors in order not to favor any particular factor specification. In addition, we evaluate the grid points only for the baseline model containing the full set of predictors, resulting in three sets of forgetting factors per model class. This should produce a parameter specification that is somewhat appropriate for all variables, while reducing the computational demand at the same time. Based on real-time data, the grid search is repeated with every GDP release to allow the forgetting factors to adjust to the economic environment. Given that our main interest is in point nowcasts for GDP growth, we then adopt the grid point that minimizes the average mean absolute deviation (MAD) over the nowcast horizons.[8] To explore the extent to which the forecast performance results are robust to the choice of optimization criterion, we report the full set of results based on "RMSE optimal" forgetting factors in the Online Appendix. Generally, the results are qualitatively and quantitatively similar.

The forgetting factor γ that governs the model switching rate is not evaluated on a grid, and it is set to the value 0.9, which implies that the forecast performance one year ago receives 65% of the weight. We set this as a slightly more aggressive value than Koop and Korobilis (2014) for two simple reasons. First, with increasing length of the forecast horizon, more time has to pass until a forecast can be evaluated and used to update the models' weights.

[8] For brevity, these results are not reported but are available from the authors upon request.

Second, the publication lags of the individual variables do not only have to be taken into account when generating the forecasts but also have to be considered when calculating the weights in real time. Since US GDP has a publication lag of around 30 days, this also delays the evaluations of the forecasts for the preceding quarter that are needed to update the DMA weights. This additionally adds considerable update lag. Our γ specification is thus chosen to counteract these two effects.

3.3. Forecast evaluation results

We conduct a recursive out-of-sample forecasting exercise to predict quarterly US GDP growth in real time. To this end, we utilize historical real-time data vintages from the FRED-MD database and quarterly US GDP from ALFRED.[9] Whenever we generate a forecast, our model thus entirely relies on information that would have been available to forecasters at the time. While our initial estimation sample spans the period from 1975M01 to 1999M12, we compute recursive forecasts for the period 2000M01 to 2021M09. Due to the Covid-19 pandemic, main economic indicators such as GDP have lately been characterized by unusually large fluctuations. Naturally, recent forecasting literature has been interested in exploring modeling approaches that account for these extreme outliers and help robustify the performance of forecasting models in abnormal times more generally (see e.g. Carriero et al., 2021, 2022; Lenza & Primiceri, 2020; Schorfheide & Song, 2021). In this context, we consider four nested model classes—MF-DFM, TVP-MF-DFM, MF-DFM-SV, and TVP-MF-DFM-SV—and thus different combinations of time-varying parameters and stochastic volatility.[10] In addition, we compute DMA forecasts for each model specification to shed some light on the additional benefits of model averaging. To establish our pre-Covid-19 baseline results, we first evaluate our forecasts for the period from 2010M01 to 2019M12.[11] Subsequently, we extend the evaluation sample to 2021M09. The results for other subsamples and the full out-of-sample period starting in 2000M1 are reported in the Online Appendix. In each of the four model classes, the benchmark specification is a stand-alone model with the full set of indicators. In contrast, the combined forecasts are based on the entire factor model space, consisting of the 49,152 individual models for each model class.

Table 2 displays the forecast performance of our benchmark model specifications and of the DMA forecasts over the evaluation period from 2010Q1 to 2019Q4. Throughout, we use the first-release GDP data to evaluate the forecast performance. While we report the MADs for the benchmark models in the upper panel, the MADs of the DMA forecasts are given in the bottom panel.[12] Finally, the forecast horizons, h, are given as the distance to the end of the target quarter measured in months. Horizons $h = 8, 7, 6$, $h = 5, 4, 3$, and $h = 2, 1, 0$ correspond to the two-quarter-ahead forecasts, the one-quarter-ahead forecasts, and the nowcasts for the ongoing quarter that are generated in the first, second, and last month of the relevant quarter, respectively.[13]

Focusing on the stand-alone model specifications in the upper panel, compared to the constant-parameter model (MF-DFM), only the model variant with stochastic volatility (MF-DFM-SV) offers slight forecast performance gains. In contrast, the model with time-varying parameters (TVP-MF-DFM) and the variant with time-varying parameters and stochastic volatility (TVP-MF-DFM-SV) suffer performance losses for all forecast horizons. These findings are similar to those obtained by Pettenuzzo and Timmermann (2017) for point forecasts. Shifting attention to the bottom panel, we find that model averaging improves the forecast performance of all models. The MF-DFM-SV (DMA) slightly improves upon the forecasting performance of the MF-DFM-SV for shorter forecasting horizons and the nowcasting horizons. In addition, the TVP-MF-DFM (DMA) and the TVP-MF-DFM-SV (DMA) improve upon the TVP-MF-DFM and the TVP-MF-DFM-SV substantially, also leading to marginal forecast performance gains vis-à-vis the MF-DFM for selected forecast horizons.[14]

After establishing our baseline results, we compare our models' performance to competitive external benchmarks from the recent forecasting literature. To cover a range of workhorse models, we select both a DFM and a VAR with mixed-frequencies, TVP, and SV. Specifically, we adopt the dynamic factor model in Antolin-Diaz et al. (2017) adjusted for forecasting by Götz and Hauzenberger (2021). Their model is characterized by mixed frequencies, time-varying long-run trends, and stochastic volatility (MFTVTDFMSV). As a VAR benchmark, we rely on the MF-VAR model with time-varying intercept and common stochastic volatility (MFTVICSV) proposed by Götz and Hauzenberger (2021). Both benchmarks thus share important features with our proposed framework and offer a similar degree of flexibility. In addition, while our model relies on an approximate algorithm, both benchmarks are estimated with fully Bayesian MCMC schemes, to shed some light on the efficiency of our fast, approximate algorithm. The MADs of both benchmark models from 2010Q1 to 2019Q4 are presented in Table 3. Starting with the MFTVTDFMSV, we find that our baseline MF-DFM with

[9] See the FRED-MD introduced by McCracken and Ng (2016) for monthly indicators and their historical vintages. For real GDP and its real-time vintages, the reader is referred to ALFRED, Federal Reserve Bank of St. Louis.

[10] Note that we compute no-change forecasts for models allowing for TVP and/or SV, meaning that we compute the forecasts directly, conditional on the last estimate of such parameters, instead of simulating their paths.

[11] Since DFMs as a nowcasting tool were first proposed by Giannone et al. (2008), this sample also avoids concerns regarding an artificial state of the art.

[12] We focus on the MAD because it is more easily interpretable. The full set of results with the RMSE is, however, reported in the Online Appendix.

[13] For the sake of brevity, we only show a table containing the mean absolute deviations in levels. We present additional tables with MADs relative to the (selected) benchmark models in Online Appendix C. Note that the pattern described above remains unchanged.

[14] The relative forecast performance of DMA forecasts compared to their stand-alone benchmark specifications is presented in Online Appendix C.

Table 2
Mean absolute deviations until 2019Q4.

$h =$	2Q-ahead			1Q-ahead			Nowcasts		
	8	7	6	5	4	3	2	1	0
MF-DFM	0.22	0.22	0.21	0.22	0.26	0.27	0.23	0.25	0.23
TVP-MF-DFM	0.28	0.27	0.25	0.26	0.29	0.28	0.26	0.27	0.23
MF-DFM-SV	0.21	0.21	0.19	0.21	0.24	0.25	0.22	0.24	0.23
TVP-MF-DFM-SV	0.28	0.27	0.26	0.27	0.30	0.29	0.25	0.27	0.24
MF-DFM (DMA)	0.22	0.22	0.21	0.22	0.28	0.28	0.22	0.27	0.24
TVP-MF-DFM (DMA)	0.23	0.23	0.21	0.25	0.30	0.28	0.24	0.28	0.24
MF-DFM-SV (DMA)	0.23	0.22	0.20	0.21	0.23	0.24	0.21	0.23	0.22
TVP-MF-DFM-SV (DMA)	0.22	0.26	0.23	0.22	0.26	0.27	0.24	0.21	0.24

Notes: This table presents mean absolute deviations (MADs) of selected model specifications over the entire evaluation sample from 2010Q1 to 2019Q4. While the top panel denotes the MADs of the benchmark model specifications, the MADs of the combined point forecasts are given in the bottom panel. The forecast horizon h is given as the distance to the end of the reference quarter measured in months.

Table 3
External benchmark models until 2019Q4.

$h =$	2Q-ahead			1Q-ahead			Nowcasts		
	8	7	6	5	4	3	2	1	0
MFTVTDFMSV	0.21	0.25	0.23	0.23	0.26	0.27	0.26	0.26	0.27
MFTVICSV	0.24	0.23	0.23	0.23	0.24	0.26	0.21	0.19	0.21

Notes: This table presents mean absolute deviations (MADs) of the MFTVTDFMSV proposed by Antolin-Diaz et al. (2017) and the MFTVICSV proposed by Götz and Hauzenberger (2021) over the entire evaluation sample from 2010Q1 to 2019Q4. The forecast horizon h is given as the distance to the end of the reference quarter measured in months.

constant parameters, the MF-DFM-SV, and the MF-DFM-SV (DMA) produce more precise forecasts for all but the longest forecast horizon. For some forecast horizons, these gains amount to roughly 20% in relative terms. In addition, all model variants with time-varying parameters improve upon the MFTVTDFMSV for some forecast horizons. Compared to the MFTVICSV, our MF-DFM-SV and MF-DFM-SV (DMA) generate more precise forecasts for the majority of forecast horizons. While the stand-alone TVP-MF-DFM and TVP-MF-DFM-SV models fall short of the benchmark for all forecast horizons, model averaging leads to performance gains for selected forecast horizons. Again, our constant-parameter MF-DFM and MF-DFM (DMA) manage to improve on the fully flexible benchmark for some forecast horizons. Note, however, that the benchmark models have a slight advantage. In contrast to the dataset used by Götz and Hauzenberger (2021), the FRED-MD database does not contain data on the S&P 500 for the ongoing months. The benchmark models thus rely on slightly more timely information on financial variables. Overall, these results demonstrate the competitiveness of our proposed forecasting model and the robustness of our approximate estimation algorithm.

Now we extend the sample of our evaluation period to 2021Q3 and thus include the Covid-19 period. The MADs are presented in Table 4. Compared to the baseline sample, the MF-DFM-SV and the MF-DFM-SV (DMA) generate sizeable performance gains over the MF-DFM. On average, the performance gains are about 12% and reach roughly 40% in the most favorable case. In contrast to the baseline sample, however, for a few forecast horizons, the TVP and the TVP-SV model variants improve upon the constant-parameter MF-DFM. Especially for the second nowcasting horizon, all models but the MF-DFM (DMA) provide sizeable performance gains.

Overall, we can make a couple interesting observations. In stable times, SV model specifications provide minor forecast performance gains over constant-parameter models. TVP and TVP-SV models only gain for some horizons when combined with model averaging. When including the Covid-19 pandemic, SV models provide sizeable gains, but TVP-SV and to a lesser extent TVP models also gain somewhat compared to the MF-DFM. The gains thus seem mostly driven by stochastic volatility, which is in line with, e.g., Carriero et al. (2021). When paired with model averaging, forecast performance again improves for all models compared to their respective single-model variant. Given that the Covid-19 sample is rather short, Tables C.5 and C.6 in the Online Appendix repeat the exercise for the global financial crisis sample, i.e. the period from 2000Q1 to 2007Q4 and 2000Q1 to 2009Q4, respectively. Overall, the results remain qualitatively similar, supporting the robustness of our findings. To establish this robustness more broadly, Section C.3 in the Online Appendix presents the full set of results using the RMSE instead. Again, the results remain qualitatively unchanged.

In light of these results, we focus on the MF-DFM-SV (DMA) specification in the following and zoom into its evolution over the first three quarters of 2020, which were especially affected by the measures taken to contain the COVID-19 pandemic. Fig. 1 illustrates the two- and one-quarter-ahead forecasts, as well as the nowcasts, made for 2020Q1 (left), 2020Q2 (middle), and 2020Q3 (right), together with the realized quarterly GDP growth.

Starting with the first quarter, the model is not able to capture the drop in GDP growth. This is mainly due

Table 4
Mean absolute deviations until 2021Q3.

h =	2Q-ahead			1Q-ahead			nowcasts		
	8	7	6	5	4	3	2	1	0
MF-DFM	0.68	0.89	0.65	0.70	1.07	0.74	0.53	0.58	0.35
TVP-MF-DFM	1.21	0.81	0.92	1.47	1.08	0.98	0.78	0.44	0.35
MF-DFM-SV	0.65	0.68	0.63	0.65	0.79	0.76	0.51	0.38	0.36
TVP-MF-DFM-SV	0.70	0.76	0.70	0.73	0.90	0.83	0.55	0.44	0.37
MF-DFM (DMA)	0.67	1.02	0.64	0.69	1.14	0.74	0.59	0.60	0.39
TVP-MF-DFM (DMA)	0.76	0.80	0.77	0.78	0.87	0.95	0.62	0.47	0.39
MF-DFM-SV (DMA)	0.66	0.68	0.64	0.66	0.78	0.73	0.51	0.34	0.36
TVP-MF-DFM-SV (DMA)	0.65	0.71	0.67	0.69	0.78	0.76	0.54	0.36	0.39

Notes: This table presents MADs of selected model specifications over the entire evaluation sample from 2010Q1 to 2021Q3. The forecast horizon h is given as the distance to the end of the reference quarter measured in months.

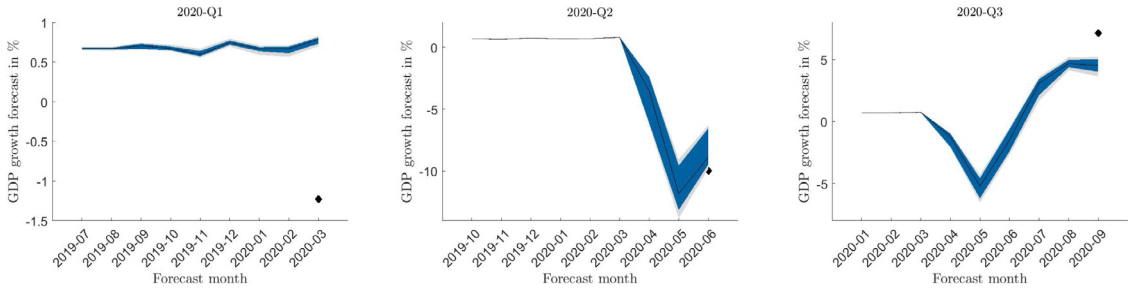

Fig. 1. Cross-sectional forecast densities for GDP growth in 2020.
Notes: This figure plots cross-sectional densities of two- and one-quarter-ahead forecasts, as well as of nowcasts, for the GDP growth in the first three quarters of 2020 obtained by the MF-DFM-SV model specifications. Blue and gray areas denote 68% and 95% bands, respectively, while the black solid line displays the median forecasts. The black dot refers to the first release of the realized quarterly GDP growth.

to the publication lag of macroeconomic indicators, such that the model only includes data up to February when the last nowcast is made at the end of March. With the inflow of data capturing the economic impact of the policy response to the pandemic, however, the model is able to adjust its nowcasts and forecasts from April on immediately. While the model only slightly underestimates the sharp drop in Q2, it falls short of the rapid recovery recorded in the following quarter. Generally, the model shows impressive flexibility in adjusting its forecasts with the real-time data flow. The adjustments amount to roughly 12 percentage points for both Q2 and Q3 and illustrate the model's ability to provide timely and reliable nowcasts and forecasts over the quarters impacted by the COVID-19 pandemic.

Note that the cross-sectional forecast densities illustrated in Fig. 1 may look somewhat narrow, especially compared to those obtained by, e.g., Bayesian and bootstrapping methods. To explore this more formally, Figure D.9 in the Online Appendix shows PITs for the period covering the Covid-19 pandemic across all forecast horizons, and indeed shows signs of underdispersion. Generally, a tight cross-sectional density indicates less disagreement between model specifications. The low level of forecast disagreement in the model space may thus be driven by, e.g., the overlap of indicators within the model space.[15]

[15] An additional source of underdispersion might originate from parameter uncertainty (especially with regard to the forgetting factors) being largely neglected in our approach.

3.4. Behind the scene of the forecast density

Integrating the stand-alone benchmark specifications into the dynamic model averaging framework enables us to observe the cross-sectional density around point forecasts for US GDP growth across the model space. In this context, it is worth noting that the cross-sectional densities are not restricted to be Gaussian and can thus assume very flexible shapes. We start with the entire cross-sectional density and then zoom into specific percentile ranges to shed light on the indicators that drive nowcasts and forecasts in the left tail of the distribution.

Fig. 2 illustrates cross-sectional densities over the 49,152 MF-DFM-SV specifications for the three nowcast horizons, $h = 2, 1, 0$, for the ongoing quarter from 2000Q3 to 2021Q3. While the black boxes show the range between the 25th and 75th percentiles, the vertical lines point to the density outside this range. Overall, the estimated cross-sectional densities are able to demonstrate forecast uncertainty reasonably well. For instance, in more turbulent times, such as during recessions and the pandemic, one can see that the variance of the forecast distribution increases drastically, and with it forecast disagreement. In contrast, between 2011 and 2019, where the US economy was located on a relatively stable growth path, the forecast distributions become heavily concentrated, and the models point in very similar directions. On the one hand, this shows that the forecast distributions from our model pick up changes in the economic

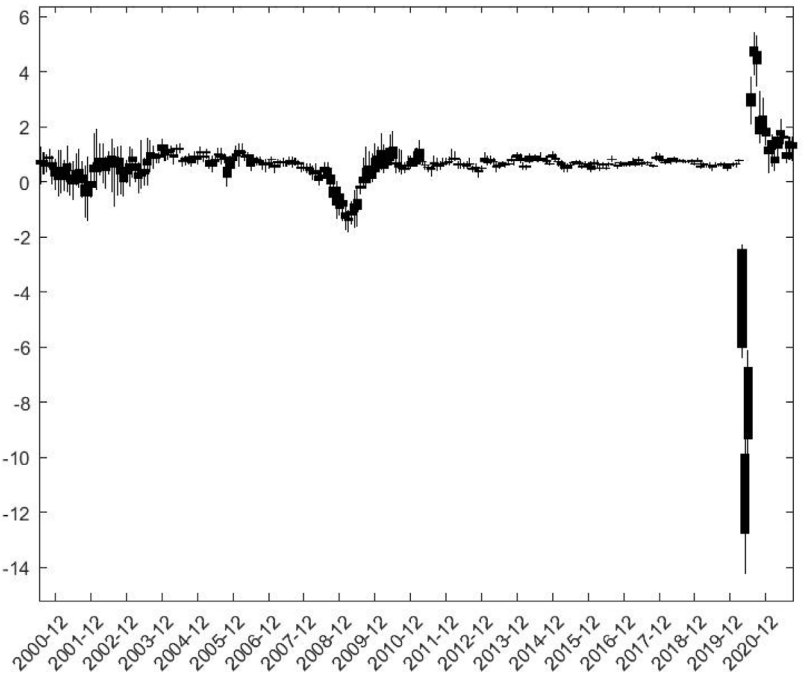

Fig. 2. Cross-sectional nowcast densities.
Notes: This figure displays the entire cross-sectional density over the each of three nowcast horizons generated by the 49,152 MF-DFM-SV specifications at $h = 2, 1, 0$ for the ongoing quarter over the evaluation sample.

conditions relatively quickly and react in a way that is reasonable. On the other hand, the changes in forecast disagreement might be useful for determining the risk that surrounds the economy at a given point in time and thus for providing insightful information to forecasters and policymakers.

After having a close look at the entire cross-sectional forecast density of the MF-DFM-SV specifications, we shift our focus to the inner workings of our suggested framework. Since factor models are based on latent and, hence, unobserved components, they are typically hard to interpret. Usually, it remains unclear why a factor model behaves a certain way or, in the context of forecasting, how forecasts are formed based on the model ingredients. As Koop and Korobilis (2011, 2014) show, one can use the updated individual model probabilities to calculate the importance of the different variables over time through the lens of DMA forecasts. Again, we extend the general framework by allowing Eq. (15) to depend on the forecast horizon:

$$\pi_{t,h}^{x_i} = \sum_{k \subset x_i} \pi_{t|t,j,h} \tag{15}$$

Fig. 3 illustrates the average importance of each additional variable in the three nowcasting horizons.[16] In the early 2000s, predictors such as housing, retail sales, 10Y treasury yields, the S&P500, and oil prices, seem to contribute relatively more to the DMA nowcasts compared to other indicators that receive rather balanced weights of medium magnitude. However, this picture is disrupted with the onset of the subprime mortgage crisis and the Great Recession. While most variables suddenly lose most of their weight, capacity utilization, initial claims, retail sales, new orders, and (to a lesser extent) the S&P 500 experience increases in their relative weights and thus their contribution to the DMA nowcasts. With the onset of the Covid-19 pandemic, the picture again changes and forecasts are predominantly driven by retail sales, new orders, consumer sentiment, and interest rates. In Online Appendix A, we also compute the expected model size, following Koop and Korobilis (2011, 2014). The chart shows considerable dynamics and indicates that neither corner solution with either all additional variables or no additional variables prevails. In conjunction with the heatmap, this shows that the additional variables are able to provide information that is not already conveyed by the five indicators that are fixed. Interested readers may also note that our measure of variable importance can easily be constructed for subspaces of the model space by summing up and re-scaling the weights appropriately. For illustrative purposes, we thus provide a heatmap that shows the average importance of the groups of variables as defined in Table 1 (see Figure D.6 in Online Appendix D). In addition, for readers more interested in coincidence indicators, Online Appendix D provides a heatmap constructed based on the weights of the models with only a single factor. In both cases, the overall results remain qualitatively similar.

[16] Figures showing the average importance of selected indicators over one- and two-quarter-ahead forecast horizons for the same period of time can be found in Online Appendix D.

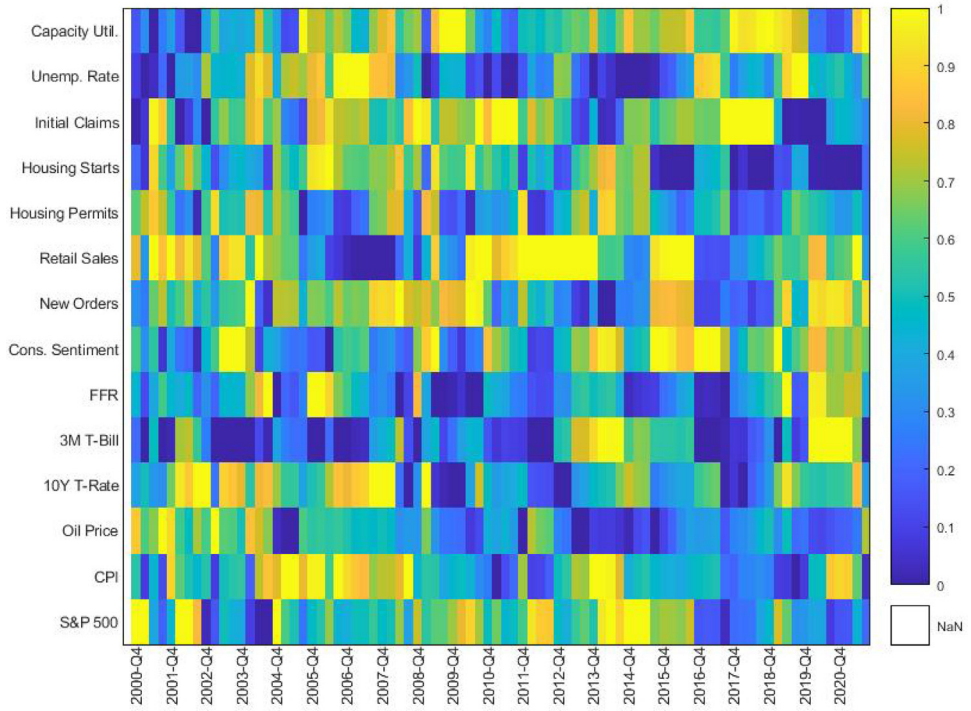

Fig. 3. Indicator heatmap for nowcast horizons.
Notes: This figure displays the average importance of the additional indicators for the three nowcast horizons in our dataset throughout the evaluation period. The heatmap is column-scaled, meaning that for each target quarter, the most (least) important indicator receives the hottest [yellow] (coldest [blue]) color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

After having looked at the entire forecast density, we now zoom into the bottom percentiles in order to highlight the indicators driving downside risks to US GDP growth forecasts.[17] Instead of estimating the magnitude of downside risks to GDP growth directly, as is mostly done in existing studies, we extract this information from our cross-sectional forecast densities. In particular, we collect the individual model forecasts that fall within a given percentile of the cross-sectional forecast distribution and then re-scale our measure of variable importance accordingly. This way, our approach can deliver model-consistent forecasts for a given percentile of any indicator included in the dataset and is able to identify the drivers of such forecasts at every forecast horizon. Fig. 4 visualizes the average importance of selected indicators in the bottom 5% of the density for the three nowcast horizons for US GDP growth.[18]

Starting with the early 2000s, the S&P 500 seems to be the most important predictor of risks to GDP nowcasts, regaining importance during the Great Recession and the second quarter of 2020. In 2002 and 2003, the oil price emerges as the most important driver of downside risks, coinciding with two major oil supply disruptions sparked by civil unrest in Venezuela and the 2003 Iraq war (see Baumeister & Kilian, 2016). Around the Great Recession, housing market and labor market indicators, the S&P 500, consumer sentiment, retail sales, and the 3M T-bill rate experience sudden spikes in importance. In contrast, during the Covid-19 pandemic, the left tail of the forecast distribution is mostly driven by the federal funds rate, the 3M T-bill rate, CPI, consumer sentiment, new orders, and retail sales. Finally, Figures D.4 and D.5 in Online Appendix D show that the federal funds rate and, to a lesser extent, the 3M Treasury bills gain importance for forecasting the left tail of GDP growth during the periods from 2005–2007 and 2016–2019, which coincide with recent interest rate increases in the US. Note that these intuitive results are not driven by any structure that we impose on the model. Rather, they seem to result naturally from the cross-sectional forecast distributions across models and hence provide tentative evidence for the validity of the modeling approach.

A few words of caution are in order, however. Although it is tempting to infer a causal relationship, especially when the changes in the models' weights seem reasonable, one must not do so. First of all, correlation structures between predictors can change over time. This holds true for indicators within and without the dataset. Rather than being important per se at a given point in

[17] The reader is referred to Adrian et al. (2019, 2018), Brownlees and Souza (2020), Figueres and Jarociński (2020), IMF (2017), Loria et al. (2019), Prasad et al. (2019), and Plagborg-Møller et al. (2020), among others, for the growing literature on downside risks to GDP growth, which is also referred to as growth at risk and vulnerable growth.

[18] Figures D.4 and D.5 in Online Appendix D show the average importance of selected indicators at the bottom 5% of the density over the one- and two-quarter-ahead forecast horizons for the same period of time.

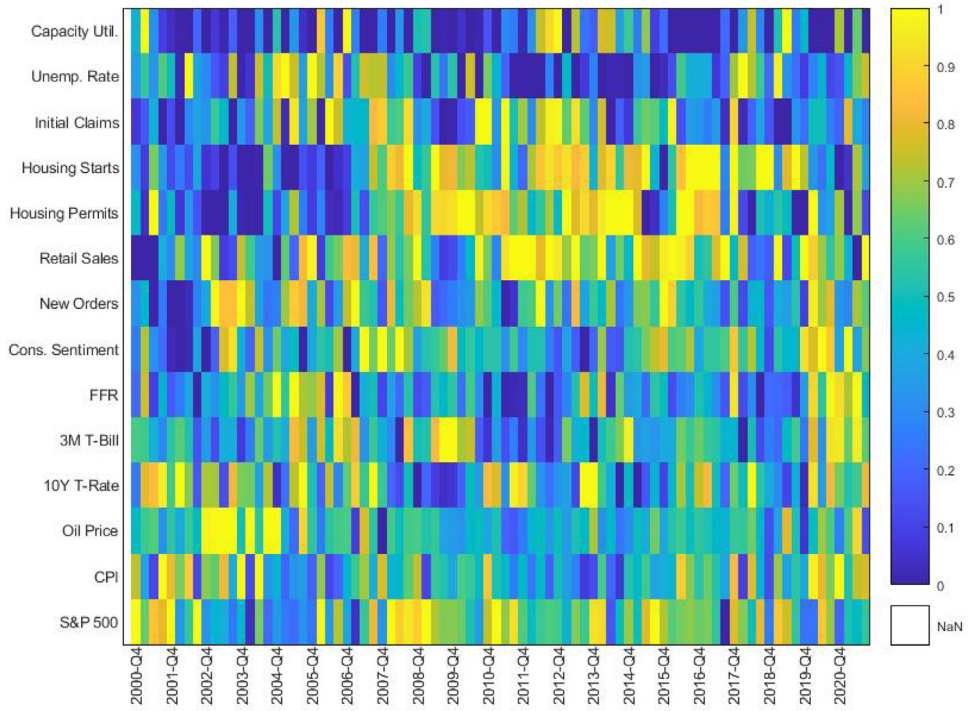

Fig. 4. Indicator heatmap at the bottom 5% of the cross-sectional density for nowcast horizons.
Notes: This figure displays the average importance of the additional indicators for the three nowcast horizons at the bottom 5% of the cross-sectional density in our dataset throughout the evaluation period. The heatmap is column-scaled, meaning that for each target quarter the most (least) important indicator receives the hottest [yellow] (coldest [blue]) color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

time, a certain indicator might receive more weight because it proxies an actually important unobserved predictor. The patterns that emerge might thus be spurious. Second, the weights used to produce these statistics are generated conditionally on the forgetting factor γ, which governs the dynamics of model switching. An increasingly lower γ implies faster model switching, which is likely to produce more erratic behavior in the expected model size and a more blurry heatmap. In the extreme case, the simple average, the heatmap will only feature one single color, as the weights are identical throughout time and across variables. These figures are thus to be understood as a tool that sheds light on how a certain forecast emerges. They provide the means to observe which and how many variables influence the forecast the most from the perspective of the model. This in turn allows the forecaster to conduct reality checks to assess whether the model places high weights on indicators that truly seem to be important at the time. Given the usual black-box character of factor models and forecasts generated by their means, this provides additional transparency to the forecaster.

Finally, our large factor model space enables us to extract recession signals from the cross-sectional forecast density. Specifically, we compute the percentage of model specifications predicting non-positive GDP growth for the ongoing quarter to obtain real-time recession probabilities on a monthly basis. Fig. 5 illustrates recession probabilities from the MF-DFM-SV specifications.

During the considered period the US economy exhibited three recessions announced by the NBER's Business Cycle Dating Committee. These periods are: March to November 2001, December 2007 to June 2009, and February 2020 to April 2020, from peak to trough.[19] Overall, Fig. 5 shows that the estimated real-time recession probabilities match up reasonably well with historical business-cycle turning points (and related recessions). This also lends support to recent studies, such as Stock and Watson (2014) and Eraslan and Nöller (2020), emphasizing the importance of large cross-sectional information for dating business-cycle turning points.

4. Concluding remarks

In this paper, we proposed a novel TVP-MF-DFM-SV-DMA nowcasting model that can efficiently deal with the characteristics of real-time data flow, as well as parameter instability and time-varying volatility. Moreover, we developed an algorithm optimized for fast estimation that allowed us to integrate our TVP-MF-DFM-SV into a dynamic model averaging framework. This enabled us to generate forecast densities based on a large model space which we used to obtain point forecasts and to shed light

[19] The Committee's announcements can be found at https://www.nber.org/research/business-cycle-dating/business-cycle-dating-committee-announcements.

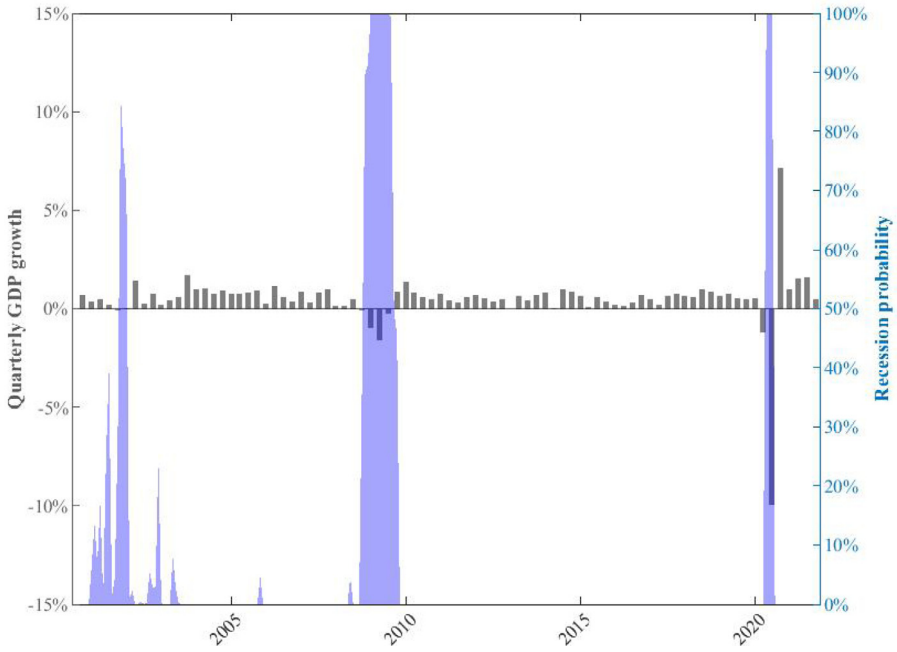

Fig. 5. Recession probabilities and GDP growth.
Notes: This figure displays quarterly GDP growth (in logs, left axis) and monthly real-time recession probabilities (blue areas, right axis) calculated as the percentage of model specifications predicting non-positive growth for the ongoing quarter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on the changing drivers of (downside risks to) US GDP growth forecasts.

We put our suggested framework to the test in a recursive out-of-sample real-time forecasting exercise. Our results revealed that SV models provide slight forecast performance gains compared to the constant-parameter model in stable times. Such gains remain absent for TVP model variants. After we established the performance of our model and the efficiency of our algorithm by comparing it to competitive literature benchmarks, we extended our sample to the Covid-19 pandemic. In line with the literature, we found that SV is the dominant driver of forecast performance gains during the pandemic. However, for some forecast horizons, we also observed performance gains derived from the TVP and TVP-SV models over the constant-parameter benchmark. In all cases, additional forecast performance gains for some forecast horizons were derived from model averaging. In this context, we also showed how the DMA methodology can be used to assess which variables are most influential for a given forecast, which provides additional transparency to the forecaster. Overall, by providing considerable improvements in forecast accuracy and additional transparency, our suggested framework is a useful complement to the forecasting toolbox of policymakers.

This paper also opens up new avenues for further research in the nowcasting and model averaging literature. Future work could continue exploring the value of the informational content of the cross-sectional forecast distributions. Moreover, our current framework could be extended to allow for different decay rates for each variable and different degrees of model change for each forecast horizon. Such extensions of our framework, however, are left for further research.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2022.07.009.

References

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*(4), 1263–1289. http://dx.doi.org/10.1257/aer.20161923.

Adrian, T., Grinberg, F., Liang, N., & Malik, S. (2018). The term structure of growth-at-risk. *IMF working paper No. 18/180*.

Aguilar, O., & West, M. (1998). Bayesian dynamic factor models and variance matrix discounting for potfolio allocation. *ISDS discussion paper 98–03*, Duke University.

Alvarez, R., Camacho, M., & Perez-Quiros, G. (2012). Finite sample performance of small versus large scale dynamic factor models. *Banco de España working paper No. 1204*.

Antolin-Diaz, J., Drechsel, T., & Petrella, I. (2017). Tracking the slowdown in long-run GDP growth. *The Review of Economics and Statistics*, *99*(2), 343–356. http://dx.doi.org/10.1162/REST_a_00646.

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Chapter 4 - now-casting and the real-time data flow. In G. Elliott, & A. Timmermann (Eds.), *Vol. 2, Handbook of economic forecasting* (pp. 195–237). Elsevier, http://dx.doi.org/10.1016/B978-0-444-53683-9.00004-9.

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting. *ECB working paper series No 1275*.

Bańbura, M., & Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary patterns of missing data. *Journal of Applied Econometrics*, *29*(1), 133–160. http://dx.doi.org/10.1002/jae.2306.

Banerjee, A., Marcellino, M., & Masten, I. (2005). Leading indicators for euro-area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics*, *67*, 785–813.

Bates, B. J., Plagborg-Møller, M., Stock, J. H., & Watson, M. W. (2013). Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics*, *177*(2), 289–304. http://dx.doi.org/10.1016/j.jeconom.2013.04.014.

Baumeister, C., & Kilian, L. (2016). Forty years of oil price fluctuations: why the price of oil may still surprise us. *Journal of Economic Perspectives*, *30*(1), 139–160. http://dx.doi.org/10.1257/jep.30.1.139.

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*, 169–194.

Brownlees, C., & Souza, A. B. (2020). Backtesting global growth-at-risk. *Journal of Monetary Economics*, http://dx.doi.org/10.1016/j.jmoneco.2020.11.003.

Byrne, J. P., Korobilis, D., & Ribeiro, P. (2018). On the sources of uncertainty in exchange rate predictability. *International Economic Review*, *59*, 329–357.

Carriero, A., Clark, T. E., Marcellino, M. G., & Mertens, E. (2021). Addressing COVID-19 outliers in BVARs with stochastic volatility. *SSRN scholarly paper ID 3816849*, Rochester, NY: Social Science Research Network.

Carriero, A., Clark, T. E., Marcellino, M. G., & Mertens, E. (2022). Measuring uncertainty and its effects in the COVID-19 era. *SSRN scholarly paper ID 3717973*, Rochester, NY: Social Science Research Network, http://dx.doi.org/10.2139/ssrn.3717973.

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, *29*(3), 327–341. http://dx.doi.org/10.1198/jbes.2010.09248.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575. http://dx.doi.org/10.1002/jae.2379.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101. http://dx.doi.org/10.1002/jae.1257.

Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, *106*(1), 157–181. http://dx.doi.org/10.1016/j.jfineco.2012.04.003.

Del Negro, M., & Otrok, C. (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. *Federal Reserve Bank of New York staff report, No. 326*.

Duarte, P., & Süssmuth, B. (2018). Implementing an approximate dynamic factor model to nowcast gdp using sensitivity analysis. *Journal of Business Cycle Research*, *14*(1), 127–141. http://dx.doi.org/10.1007/s41549-018-0026-0.

Durbin, J., & Koopman, J. (2012). *Time series analysis by state space methods*. Oxford University Press.

Elliott, G., Gargano, A., & Timmermann, A. (2013). Complete subset regressions. In *Dynamic econometric modeling and forecasting*: *Journal of Econometrics*, In *Dynamic econometric modeling and forecasting*: *177*(2), 357–373.http://dx.doi.org/10.1016/j.jeconom.2013.04.017,

Eraslan, S., & Nöller, M. (2020). Recession probabilities falling from the STARs. *Deutsche Bundesbank Discussion Paper, No 08/2020*.

Farmer, L., Schmidt, L., & Timmermann, A. (2018). Pockets of predictability. *SSRN scholarly paper ID 3152386*, Rochester, NY: Social Science Research Network, http://dx.doi.org/10.2139/ssrn.3152386.

Figueres, J. M., & Jarociński, M. (2020). Vulnerable growth in the euro area: Measuring the financial conditions. *Economics Letters*, *191*, Article 109126. http://dx.doi.org/10.1016/j.econlet.2020.109126.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*, 665–676. http://dx.doi.org/10.1016/j.jmoneco.2008.05.010.

Götz, T. B., & Hauzenberger, K. (2021). Large mixed-frequency VARs with a parsimonious time-varying parameter structure. *The Econometrics Journal*, *24*(3), 442–461. http://dx.doi.org/10.1093/ectj/utab001.

IMF (2017). *Global financial stability report: Is growth at risk?*. Washington, DC: International Monetary Fund.

Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. In *Recent advances in time series econometrics*: *Journal of Econometrics*, In *Recent advances in time series econometrics*: *178*, 352–367.http://dx.doi.org/10.1016/j.jeconom.2013.08.033,

Koop, G., & Korobilis, D. (2011). UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling*, *28*(5), 2307–2318. http://dx.doi.org/10.1016/j.econmod.2011.04.008.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*(3), 867–886.

Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, *71*, 101–116. http://dx.doi.org/10.1016/j.euroecorev.2014.07.002.

Lenza, M., & Primiceri, G. E. (2020). How to estimate a VAR after march 2020. *ECB working paper series No 2461*.

Loria, F., Matthes, C., & Zhang, D. (2019). Assessing macroeconomic tail risk. *FRB Richmond working paper No. 19-10*.

Marcellino, M., Porqueddu, M., & Venditti, F. (2016). Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics*, *34*(1), 118–127. http://dx.doi.org/10.1080/07350015.2015.1006773.

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443. http://dx.doi.org/10.1002/jae.695.

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589. http://dx.doi.org/10.1080/07350015.2015.1086655.

Muth, J. F. (1960). Optimal properties of exponentially weighted forecasts. *Journal of the American Statistical Association*, *55*(290), 299–306.

Nakajima, J., & West, M. (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics*, *31*(2), 1.

Onorante, L., & Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review*, *81*, 2–14. http://dx.doi.org/10.1016/j.euroecorev.2015.07.013.

Pettenuzzo, D., & Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, *35*(2), 183–201. http://dx.doi.org/10.1080/07350015.2015.1051183.

Plagborg-Møller, M., Reichlin, L., Ricco, G., & Hasenzagl, T. (2020). When is growth at risk? *Brookings Papers on Economic Activity*, *2020 (Spring)*, 167–229.

Prasad, M. A., Elekdag, S., Jeasakul, M. P., Lafarguette, R., Alter, M. A., Feng, A. X., & Wang, C. (2019). Growth at risk: Concept and application in imf country surveillance. *IMF working paper No. 19/36*, International Monetary Fund.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852.

Raftery, A. E., Kárný, M., & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, *52*(1), 52–66.

Rossi, B. (2013). Chapter 21 - advances in forecasting under instability. In G. Elliott, & A. Timmermann (Eds.), *Handbook of economic forecasting*: *Vol. 2*, (pp. 1203–1324). Elsevier, http://dx.doi.org/10.1016/B978-0-444-62731-5.00021-X.

Schorfheide, F., & Song, D. (2021). Real-time forecasting with a (standard) mixed-frequency var during a pandemic. In *Working paper series 29535*, National Bureau of Economic Research, http://dx.doi.org/10.3386/w29535.

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicators. In G. Moore, & K. Lahiri (Eds.), *The leading economic indicators: new approaches and forecasting records* (pp. 63–90). Cambridge University Press.

Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. http://dx.doi.org/10.1198/016214502388618960.

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. http://dx.doi.org/10.1198/073500102317351921.

Stock, J. H., & Watson, M. W. (2014). Estimating turning points using large data sets. *Journal of Econometrics*, *178*, 368–381. http://dx.doi.org/10.1016/j.jeconom.2013.08.034.

Thorsrud, L. A. (2020). Words are the new numbers: A newsy coincident index of the business cycle. *Journal of Business & Economic Statistics*, *38*(2), 393–409. http://dx.doi.org/10.1080/07350015.2018.1506344.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models*. Springer-Verlag New York, Inc..