LATVIJAS BANKA
EIROSISTĒMA

ISBN 978-9934-578-33-5

BORISS SILIVERSTOVS

GAUGING THE EFFECT OF INFLUENTIAL OBSERVATIONS ON MEASURES OF RELATIVE FORECAST ACCURACY IN A POST-COVID-19 ERA: APPLICATION TO NOWCASTING EURO AREA GDP GROWTH

WORKING PAPER

1 / 2021

Latvijas Banka
K. Valdemāra iela 2A, Riga, LV-1050
Tel.: +371 67022300 info@bank.lv
https://www.bank.lv https://www.macroeconomics.lv

CONTENT

ABBREVIATIONS

ARM – autoregressive model
CSSFED – cumulated sum of squared forecast error difference
DFM – dynamic factor model
EDC – European Debt Crisis
ESI – Economic Sentiment Indicator
EXI – intra-euro area exports of goods
EXJ – extra-euro area exports of goods
GFC – the Great Financial Crisis
GDP – gross domestic product
HMM – historical mean model
IIP – index of industrial production
IOR – industrial orders
MSFE – mean squared forecast error
NRM – normal
OOS – out-of-sample
R^2MSFE – recursive relative mean squared forecast error
R^3MSFE – recursive relative root mean squared forecast error
R^2MSFE(+R) – recursive relative mean squared forecast error (based on rearranged observations)
R^3MSFE(+R) – recursive relative root mean squared forecast error (based on rearranged observations)
rMSFE – relative mean squared forecast error
RMSFE – root mean squared forecast error
SE – standard error
SFED – squared forecast error difference

ABSTRACT

The previous research already emphasised the importance of investigating the predictive ability of econometric models *separately* during expansions and recessions (Chauvet and Potter (2013), Siliverstovs (2020), Siliverstovs and Wochner (2020)). Using the data for the pre-COVID period, it has been shown that ignoring asymmetries in a model's forecasting accuracy across the business cycle phases typically leads to a biased judgement of the model's predictive ability in each phase. In this study, we discuss the implications of data challenges posed by the COVID-19 pandemic on econometric model estimates and forecasts. Given the dramatic swings in GDP growth rates across a wide range of countries during the coronavirus pandemic, one can expect that the asymmetries in the models' predictive ability observed during the pre-COVID period will be further exacerbated in the post-COVID era. In such situations, recursive measures that dissect the models' forecasting ability observation by observation allow to gain detailed insights into the underlying causes of one model's domination over the others. In this paper, we suggest a novel metric referred to as *the recursive relative mean squared forecast error (based on rearranged observations)* or *R^2MSFE(+R)*. We show how this new metric paired with the cumulated sum of squared forecast error difference (CSSFED) of Welch and Goyal (2008) highlights significant differences in the relative forecasting ability of the dynamic factor model and naive univariate benchmark models in expansions and recessions that are typically concealed when only point estimates of relative forecast accuracy are reported.

Keywords: COVID-19, nowcasting, GDP, euro area

JEL codes: C22, C52, C53

The author is grateful to Konstantin A. Kholodilin for his helpful and constructive comments on an earlier version of this paper. Part of the material was presented at the 2nd Vienna Workshop on Forecasting. The author also thanks the participants for their helpful comments. The views are solely of the author and under no circumstances represent those of Latvijas Banka. All computations were performed in R (R Core Team, 2020).

1 INTRODUCTION

In a recent paper, Lenza and Primiceri (2020) investigate the consequences of historically unprecedented outliers in macroeconomic time series brought about by the COVID-19 pandemic for estimation of vector autoregressive (VAR) models. These outliers, unless handled properly, not only distort estimation, inference and forecasting outcomes in macroeconometric models but also have serious implications for how these models are evaluated based on their out-of-sample forecasting performance. In this study, we investigate the consequences of these outliers for absolute and relative measures of forecast accuracy and suggest a metrics to gauge their influence on the metrics used to gauge the models' relative forecasting performance.

Our contribution to the forecasting literature is motivated by the following observations based on the empirical forecasting literature:

1) different observations have different contributions to standard measures of forecast accuracy, e.g. mean squared forecast error (MSFE) (Siliverstovs (2017));

2) only a few observations may be **pivotal** in making one model preferable to another based on their relative forecast accuracy (Geweke and Amisano (2010));

3) these few observations typically occur in recessions (Siliverstovs (2020);

4) during normal times, simple univariate benchmark models are hard to beat (Chauvet and Potter (2013));

5) forecasting gains during recessions typically significantly overweigh the mediocre performance of sophisticated models during normal times. As a result, the forecasting accuracy of more sophisticated models tends to be **overstated (**Siliverstovs and Wochner (2020)).

Notwithstanding these observations, recently released research papers also continue to ignore asymmetry in the forecasting ability of the commonly used forecasting models during economic expansions and recessions and report measures of average forecast accuracy for a full period encompassing one or several recessions, including the Great Recession (Cimadomo et al. (2020)). In the best case, forecast accuracy measures are additionally reported for recessionary periods or only for the observations during the Great Recession (Delle Monache et al. (2020)). As discussed in the references above, such a practice typically results in a biased judgment artificially favouring the forecasting performance of a more sophisticated model relative to the simple benchmark models. In this study, we argue that such malicious practice cannot be continued in the presence of unprecedentedly large swings in the quarterly GDP growth commonly observed during the outbreak of the COVID-19 pandemic in the second and third quarters of 2020. The forecast errors are so large that they cannot be simply swept under the carpet as it often was the case with the observations during the previous recessions, and future forecast evaluation exercises need to be open about the excessive influence of these extreme observations on commonly used forecast accuracy measures.

The rest of the paper is organised as follows. Section 2 outlines the data on the quarterly euro area GDP growth rate. In Section 3, we present the mixed-frequency dynamic factor model used to forecast the euro area GDP growth rate and compare its

average forecasting accuracy with that of the historical mean model (HMM) for the whole out-of-sample forecast evaluation period (from the first quarter of 2006 to the third quarter of 2020) for illustrative purposes, ignoring the differences in the forecasting performance during economic expansions and the following sub-periods of economic distress: the Great Recession (from the second quarter of 2008 to the second quarter of 2009), the European Debt Crisis (from the fourth quarter of 2011 to the first quarter of 2013), and the COVID-19 pandemic (from the first quarter of 2020 to the third quarter of 2020). Section 4 contains a description of the recursive forecasting exercise. In Section 5, we present the recursive forecast evaluation metrics that we find very useful in identifying influential observations and assessing their effect on absolute and relative measures of forecast accuracy. The first one is the cumulated sum of squared forecast error difference (CSSFED) introduced in Welch and Goyal (2008), whereas the second one, referred to as recursive relative mean squared forecast error (based on rearranged observations), is for the first time introduced in this paper. For the sake of brevity, we use the abbreviation $R^2MSFE(+R)$. In Section 6, we illustrate how these two complementary forecast accuracy measures can be used to gain a thorough insight in the forecasting properties of the competing models. The first one is useful to identify influential observations measured by squared forecast error difference (SFED), whereas the second one is useful to gauge the effect of these influential observations on the relative MSFE commonly computed in the forecasting literature. In Section 7, we compare the effect on forecast accuracy of two approaches to dealing with the data challenges posed by the COVID-19 pandemic. The first approach is to continue generating forecasts based on the recursively estimated coefficients as it was done for the rest of the sample. The alternative is to generate forecasts using coefficient values frozen at their pre-COVID period estimates. The final section concludes.

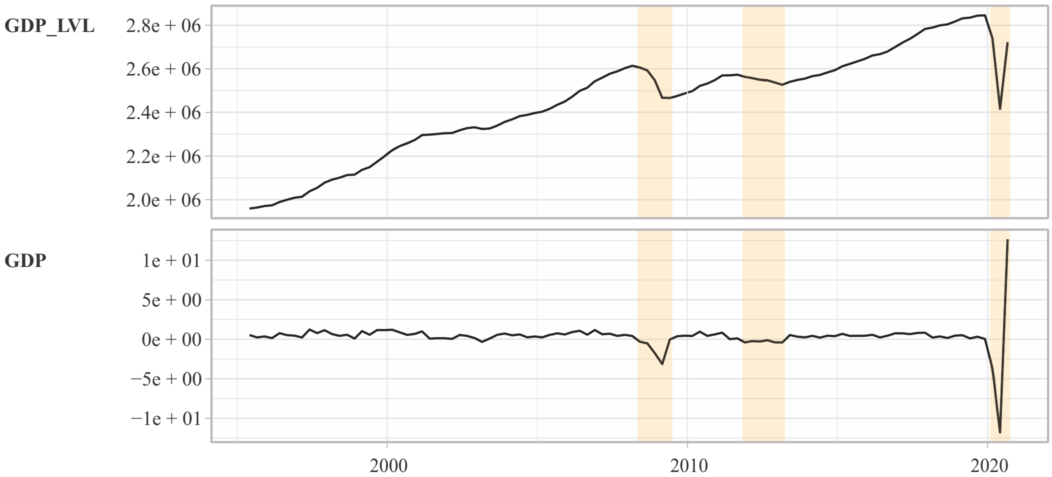
2 DATA

In this exercise, we use the data vintages downloaded from the ECB Statistical Data Warehouse and the European Commission website on 2 November 2020. Our target variable is the euro area GDP at chain-linked prices. For our out-of-sample forecasting exercise of GDP growth we use the following auxiliary monthly time series: 1) euro area industrial production excluding construction (IIP); 2) intra-euro area exports of goods (EXI); 3) extra-euro area exports of goods (EXJ); 4) euro area industrial orders (IOR); 5) euro area Economic Sentiment Index (ESI). Consequently, there are four hard and one soft indicators. The soft indicator is characterised by a shorter publication lag than its hard counterparts. The hard indicators are the same as in Perez-Quiros et al. (2020).

The time series of the euro area GDP are shown in Figure 1. The level is displayed in the upper panel and the derived quarterly growth rate is displayed in the lower panel. The COVID-19 pandemic induced unprecedented swings in the GDP dynamics. The trough reached in the level in the second quarter of 2020 is as low as the real GDP level observed in 2005. Naturally, this is reflected in the growth rates. In the first quarter of 2020, the estimated GDP growth rate was comparable to the worst drop in GDP during the Great Recession. In the second and third quarters of 2020, we observed an equally staggering fall and a subsequent rise in GDP of about 11.8 and 12.5 percentage points respectively. The sample period from the first quarter of 1995 to the third quarter of 2020 includes the three recessionary periods – the Great

Financial Crisis (GFC), the European Debt Crisis (EDC) and the COVID-19 pandemic (CVD) – shown in the figures by the shaded area.

Figure 1
Euro area GDP
(vintage from 2 November 2020; at chain-linked prices)



Naturally, the above-discussed swings in GDP are also reflected in the monthly indicators. This fact, combined with shorter publication lags of these indicators, can be capitalised upon when nowcasting GDP growth out of sample. In Figure 2, we show the levels of five monthly indicators, whereas the monthly growth rates of the four hard indicators are displayed in Figure 3.

3 DYNAMIC FACTOR MODEL

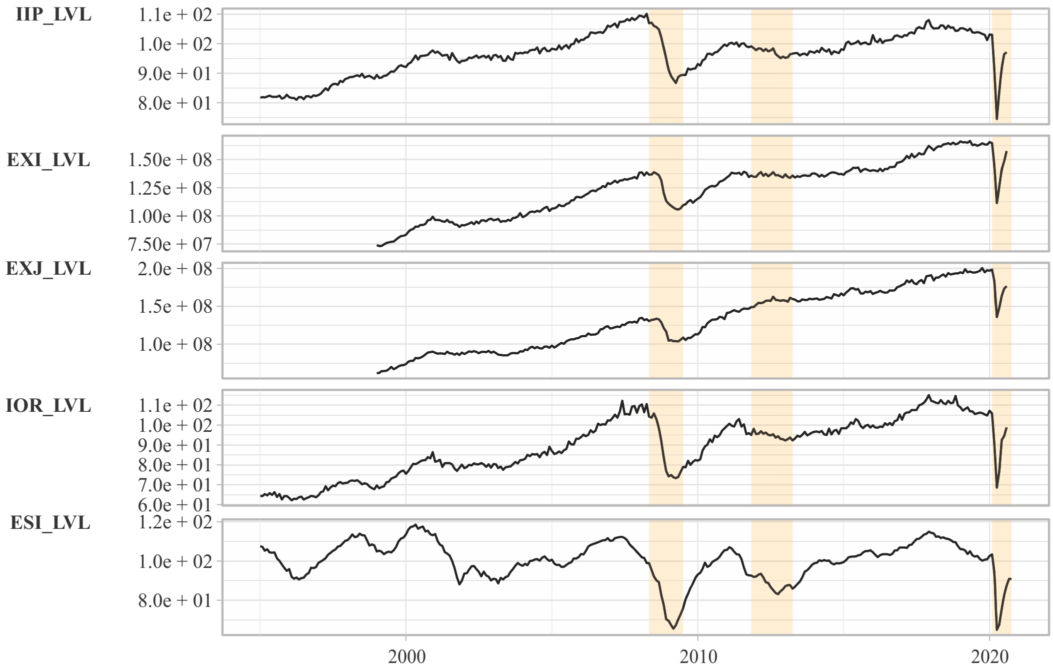
The specification of the econometric model for producing nowcasts of GDP growth closely follows Camacho and Perez-Quiros (2010). Camacho and Perez-Quiros (2010) suggest a dynamic factor model that combines both data sampled at quarterly y_t and monthly $z_t = (z_t^{IIP}, z_t^{EXI}, z_t^{EXJ}, z_t^{IOR}, z_t^{ESI})'$ frequencies. An additional feature of the model is that it allows data entry in different transformations, i.e. it eclectically combines quarterly growth of GDP, monthly growth rates of hard indicators and levels of survey data. The model is cast into a state-space form and the Kalman filter is used to estimate model parameters and make out-of-sample forecasts.

We define a common latent factor f_t driving an equally unobserved monthly growth rate of euro area GDP y_t^*:

$$y_t^* = \beta f_t + u_t.$$

As noted in Mariano and Murasawa (2003), the observed quarterly GDP growth rate can be approximated by a deterministic linear combination of latent monthly growth rates $y_t = 1/3 y_t^* + 2/3 y_{t-1}^* + y_{t-2}^* + 2/3 y_{t-3}^* + 1/3 y_{t-4}^*$. Then a combination of these two equations links the GDP growth with the latent factor:

Figure 2
Monthly indicators
(levels; vintage from 2 November 2020)



$$y_t = \beta\left(\frac{1}{3}f_t + \frac{2}{3}f_{t-1} + f_{t-2} + \frac{2}{3}f_{t-3} + \frac{1}{3}f_{t-4}\right) + \frac{1}{3}u_t + \frac{2}{3}u_{t-1} + u_{t-2} + \frac{2}{3}u_{t-3} + \frac{1}{3}u_{t-4}.$$

Next, we specify the following model equations, directly linking the observed monthly growth rates of hard indicators to the latent factor:

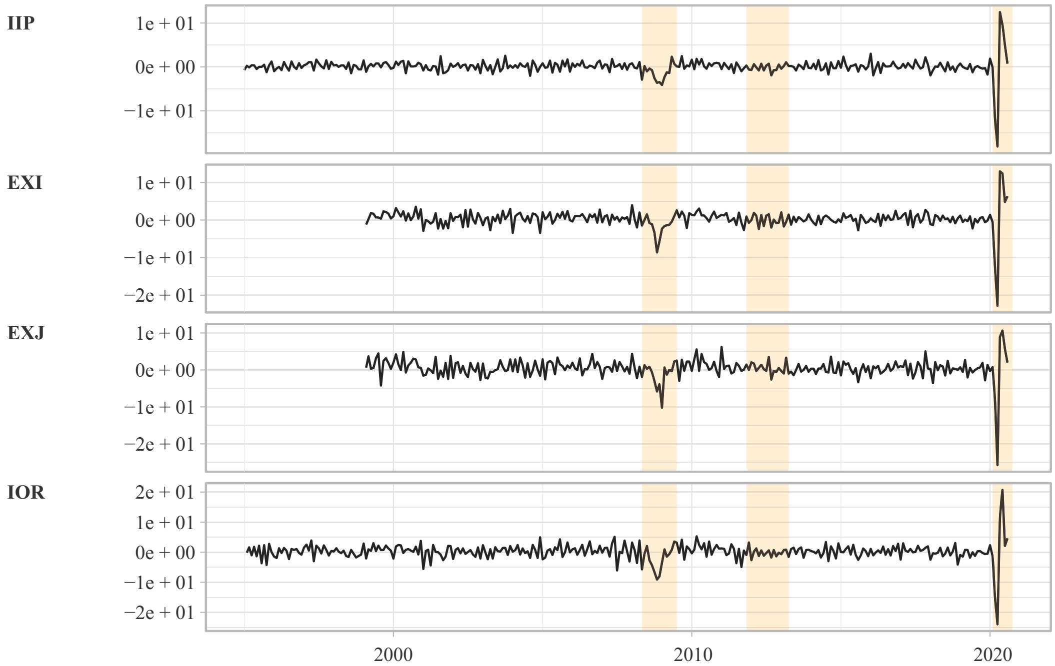
$$z_t^i = \lambda^i f_t + v_t^i, \quad \text{with} \quad i = IIP, EXI, EXJ, IOR.$$

Regarding the linkage between the observed survey variable and the latent factor, we follow Camacho and Perez-Quiros (2010):

$$z_t^{ESI} = \lambda^{ESI}\sum_{j=0}^{11} f_t + v_t^{ESI}.$$

In order to account for the more persistent nature of the ESI variable, Camacho and Perez-Quiros (2010) suggest to approximate those by means of summation over twelve most recent values of the latent factor.

Figure 3
Monthly indicators
(monthly growth; vintage from 2 November 2020)



The model specification is completed by stipulating that the latent factor model follows an AR(1) process:

$$f_t = \phi f_{t-1} + \varepsilon_t$$

and the assumption that all error terms in the model are white noise variables, i.e. $u_t \sim N_{iid}(0, \sigma_u^2)$, $\varepsilon_t \sim N_{iid}(0, \sigma_f^2)$, $v_t^i \sim N_{iid}(0, \sigma_i^2)$ with $i \in \{IIP, EXI, EXJ, IOR, ESI\}$. In making this simplifying assumption, we depart from the model specification of Camacho and Perez-Quiros (2010) where the idiosyncratic components were also allowed to follow a higher-order autoregressive process. Notwithstanding this, since the chosen parsimonious model specification fully serves the main purpose of this paper, we decided not to over-complicate the model specification beyond what is strictly necessary and sufficient to make our point (see Antolin-Diaz et al. (2017), Antolin-Diaz et al. (2020), and Delle Monache et al. (2020) for possible extensions of our simple model).

The following coefficient estimates are of our main interest. The estimate of factor loading coefficient to the observed quarterly GDP growth β as well as loading coefficients to each of the monthly auxiliary variables λ^i with $i \in \{IIP, EXI, EXJ, IOR, ESI\}$. Since all the monthly variables are pro-cyclical business cycle indicators, we expect all factor loading coefficient estimates to be positive and statistically significant $\hat{\beta} > 0$ and $\hat{\lambda}^i > 0$ with $i \in \{IIP, EXI, EXJ, IOR, ESI\}$. Last but not least, we impose the following identifying restriction: $\sigma_f^2 = 1$.

4 FORECASTING FRAMEWORK

We conduct the forecasting exercise using pseudo-real time data, which is not that uncommon in the forecasting literature (Marcellino and Schumacher (2010)). The data set was downloaded on 2 November 2020. At this point, the first estimate of the euro area GDP for the third quarter of 2020 was already available. This implies that our forecast evaluation sample ends with the third quarter of 2020. The first quarter for which we make nowcast is the first quarter of 2006, i.e. it precedes the Great Recession. All in all, we have 59 observations out of sample.

For each quarter in this out-of-sample forecast evaluation sample, we specify the forecast origin in the beginning of the first month of the following quarter. At this forecast origin, we have maximum information about the target quarter from the auxiliary monthly variables before the official release of the first estimate of GDP for this quarter. The four hard monthly indicators extend until the first two months of the target quarter, whereas the soft survey variable has observations for all three months in the quarter of interest. In line with the rest of the literature on mixed-frequency data sets (involving both quarterly and monthly data), we assume that the quarterly GDP observations are placed in the last month of each quarter.

Table 1
Vintage at forecast origin for nowcasting the third quarter of 2020

Date	GDP	IIP	EXI	EXJ	IOR	ESI
2020 I	NA	1.9	1.4	-0.3	2.3	102.6
II	NA	–0.1	–0.5	0.7	–1.3	103.4
III	–3.7	–11.7	–12.4	–7.8	–14.8	94.1
IV	NA	–18.1	–22.8	–25.7	–24.0	64.9
V	NA	12.5	13.0	9.0	11.8	67.5
VI	–11.8	9.5	12.4	10.6	20.7	75.8
VII	NA	5.0	4.8	5.7	2.1	82.4
VIII	NA	0.7	6.3	2.0	4.6	87.5
IX	NA	NA	NA	NA	NA	90.9

Note: The variables are transformed: quarterly growth for GDP, monthly growth for IIP, EXI, EXJ, IOR and levels for ESI.

These different publication lags result in an unbalanced data set with a ragged bottom edge. When constructing pseudo-real time vintages, we truncate the initial data set to preserve this pattern of the ragged edge at each forecast origin. An example of this ragged-edge data set used to generate the forecast for the third quarter of 2020 is shown in Table 1. Each pseudo-real time data vintage starts in the first quarter of 1995, dictated by the availability of the GDP time series.

The forecasts are made recursively by using an expanding window of observations. The first sample used for the estimation of model parameters and making a forecast of GDP growth in the first quarter of 2006 respectively is from the first quarter of 1995 to the first quarter of 2006, with all the missing values present as discussed above. The GDP forecast for the second quarter of 2006 is made using the sample from the first quarter of 1995 to the second quarter of 2006, and so on up to the forecast for the third quarter of 2020. The GDP forecast is taken from the filtering step output of the Kalman filter routine each time reported for the last month of the quarter of interest.

5 EVALUATION METRICS

In this section, we present the forecast accuracy evaluation metrics that we use in our study. A standard point estimate of the average forecasting performance over some period of time that is P observations long is the mean squared forecast error (MSFE):

$$MSFE_i^P = \frac{1}{P}\sum_{t=1}^{P}\left(y_t - \hat{y}_{i,t}\right)^2 = \frac{1}{P}\sum_{t=1}^{P}\hat{e}_{i,t}^2.$$

A pair of models is compared using the relative MSFE:

$$rMSFE_{1/2}^P = \frac{MSFE_1^P}{MSFE_2^P} - 1,$$

indicating a relative improvement in MSFE achieved by one model in comparison with the other model. Positive values of $rMSFE_{1/2}$ indicate that model 1 is, on average, less accurate than model 2 based on the evidence from all P observations. $rMSFE_{1/2}$ is also a point estimate of the relative average forecasting performance of the competing models.

A conceptually different but closely related metrics for forecast accuracy evaluation was suggested in Welch and Goyal (2008), a so-called cumulative sum of squared forecast error difference (CSSFED):

$$CSSFED_{1/2,t}^{[\underline{\tau},\overline{\tau}]} = \sum_{t=\underline{\tau}}^{\overline{\tau}}(\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2).$$

The CSSFED can be plotted against time over the out-of-sample (OOS) forecast evaluation period $[\underline{\tau},\overline{\tau}]$ and at each point of time t the cumulated difference between squared forecast errors $\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2$ computed up to and including observation at t can be visually assessed. In case of no systematic differences in forecast accuracy, the CSSFED will display minor fluctuations around some levels. If there is a noticeable systematic difference in forecast accuracy between models, e.g. $(\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2) > 0$ for a lion share of observations t in the OOS forecast evaluation period, the CSSFED will display upward trending dynamics. If the opposite is the case, then the CSSFED will display a tendency to move downwards. In cases when for an observation t^* the corresponding value of $SFED_{t^*}$ substantially deviates from the SFED values typically observed for the rest of the observations, the CSSFED sequence will be characterised by an abrupt jump in either direction depending on the sign of $SFED_{t^*}$. By observing such jumps in $CSSFED_{1/2,t}^{[\underline{\tau},\overline{\tau}]}$, one can easily identify the observations with the largest contribution to the reported differences in forecast accuracy.

Such an observation-by-observation dissection of the models' relative forecast accuracy allows to discover the genuine sources of the forecast accuracy domination of one model over the other that are usually hidden in measures of relative forecast accuracy based on averages (Siliverstovs (2017)). For example, it is easy to see whether one model produces lower MSFE because it consistently produces lower squared forecast errors or because of few noticeable observations for which the squared forecast error differential is much larger compared to that observed for most of the remaining observations (see Siliverstovs (2020) for application of the CSSFED when evaluating the forecasting performance of competing models for US GDP growth). The usefulness of recursive measures of relative forecasting accuracy, though in the Bayesian context, has been highlighted earlier in Geweke and Amisano (2010).

The CSSFED is a recursively computed sequence of cumulated differences of squared forecast errors evaluated at each point $t \in [\underline{\tau}, \overline{\tau}]$. Denoting the corresponding length of the sample $P_t^{[\underline{\tau},\overline{\tau}]} = t - \underline{\tau} + 1$ for $\forall t \in [\underline{\tau}, \overline{\tau}]$ the CSSFED can be represented as the pointwise sequence of **arithmetical** differences in model-specific MSFEs scaled by the corresponding number of observations $P_t^{[\underline{\tau},\overline{\tau}]}$:

$$CSSFED_{1/2,t}^{[\underline{\tau},\overline{\tau}]} = \sum_{t=\underline{\tau}}^{\overline{\tau}} (\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2) = \sum_{t=\underline{\tau}}^{\overline{\tau}} \hat{e}_{1,t}^2 - \sum_{t=\underline{\tau}}^{\overline{\tau}} \hat{e}_{2,t}^2$$
$$= (MSFE_{1,t}^{[\underline{\tau},\overline{\tau}]} - MSFE_{2,t}^{[\underline{\tau},\overline{\tau}]}) * P_t^{[\underline{\tau},\overline{\tau}]}.$$

A related question can be asked on whether it is possible to construct a recursively evaluated measure of $relative$ forecast accuracy which could show that the relative forecast accuracy evolves as more observations are added to the out-of-sample forecast evaluation period. It turns out that such measure, *recursive relative MSFE* (R^2MSFE), is easy to compute. Adopting Equation (1) to recursive estimation

$$R^2MSFE_{1/2,t}^{[\underline{\tau},\overline{\tau}]} = \frac{MSFE_{1,t}^{[\underline{\tau},\overline{\tau}]}}{MSFE_{2,t}^{[\underline{\tau},\overline{\tau}]}} - 1 = \frac{(MSFE_{1,t}^{[\underline{\tau},\overline{\tau}]} - MSFE_{2,t}^{[\underline{\tau},\overline{\tau}]}) * P_t^{[\underline{\tau},\overline{\tau}]}}{MSFE_{2,t}^{[\underline{\tau},\overline{\tau}]} * P_t^{[\underline{\tau},\overline{\tau}]}}$$

results in

$$R^2MSFE_{1/2,t}^{[\underline{\tau},\overline{\tau}]} = \frac{CSSFED_{1/2,t}^{[\underline{\tau},\overline{\tau}]}}{CSSFE_{2,t}^{[\underline{\tau},\overline{\tau}]}} = \frac{\sum_{t=\underline{\tau}}^{\overline{\tau}}(\hat{e}_{1,t}^2 - \hat{e}_{2,t}^2)}{\sum_{t=\underline{\tau}}^{\overline{\tau}} \hat{e}_{2,t}^2} \qquad (1).$$

Despite its computational simplicity, this measure in Equation (1) is rarely seen in the forecasting literature (e.g. see Baumeister and Guerin (2020) for a recent application). It has the property that subsequent values are heavily influenced by earlier values, especially if some of them are very large. In such cases, adding observations to the OOS that are characterised by relatively small squared forecast error differentials results in little effect on R^2MSFE and hence little variation in this sequence. Therefore, the leverage of the few but most influential observations on the model ranking based on the recursive relative MSFEs is difficult to decipher. However, similarly to the CSSFED sequence, plotting the R^2MSFE against time is helpful to visually identify the timing of these pivotal observations.

Another common approach to capture time-varying relative forecasting ability is to report rMSFEs computed over a fixed-width rolling window. This way the influence of very distant large observations on the local measures of relative forecast accuracy is eliminated at the moment they fall out of the rolling window span. As a result, the reported sequence of the relative forecast accuracy displays more variation over time. These local measures based on averages computed over rolling samples are subject to the same critique as the global measures computed over the whole sample in the sense that detection of the contribution of each data point is similarly obscure and the width of the rolling window is an arbitrary chosen hyper-parameter. For example, Banbura and Bobeica (2020) rely on a rolling window width of 20 quarters when evaluating the inflation forecasting performance of different Phillips curve models for the euro area.

We argue that a simple operation, namely, rearranging the squared forecast error differences according to their absolute magnitude in an ascending order may make this recursive metrics more appealing in the applied work. The rearrangement operation particularly allows us to extract the effect of the few pivotal observations with the largest squared forecast error differences on the aggregate relative forecast accuracy of the competing models. We refer to this new measure as the recursive relative mean squared forecast error (based on rearranged observations) or $R^2MSFE(+R)$.

The $R^2MSFE(+R)$ is constructed as follows. First, we rearrange the squared forecast error difference (SFED) and, correspondingly, the squared forecast errors (SFE) for each model in an ascending order by its absolute magnitude: $|e_{1,j}^2 - e_{2,j}^2|$ with $j_i < j_k$ whenever $|e_{1,j_i}^2 - e_{2,j_i}^2| < |e_{1,j_k}^2 - e_{2,j_k}^2|$. Then the $R^2MSFE(+R)$ is obtained in Equation (2) by applying a formula to these rearranged sequences of $SFED_{1/2}$ and SFE_2:

$$R^2MSFE(+R)_{1/2}^{[j_1,\dots,j_i,\dots,j_P]} = \frac{\sum_{j=j_1}^{j_P}(e_{1,j}^2 - e_{2,j}^2)}{\sum_{j=j_1}^{j_T} e_{2,j}^2} = \frac{\sum_{j=j_1}^{j_P} e_{1,j}^2}{\sum_{j=j_1}^{j_P} e_{2,j}^2} - 1 \quad (2),$$

with P indicating the total number of observations in the OOS, $P = P_{\underline{\tau}}^{[\underline{\tau},\overline{\tau}]} = \overline{\tau} - \underline{\tau} + 1$.

Observe that if the target variable is characterised by well-defined distinct states (like GDP with recession and expansion phases of the business cycle), then this rearranging operation can be done for each distinct state separately. In the empirical illustration below, we will show how this additional detail helps in highlighting the contrast in the relative forecasting performance between the business cycle phases.

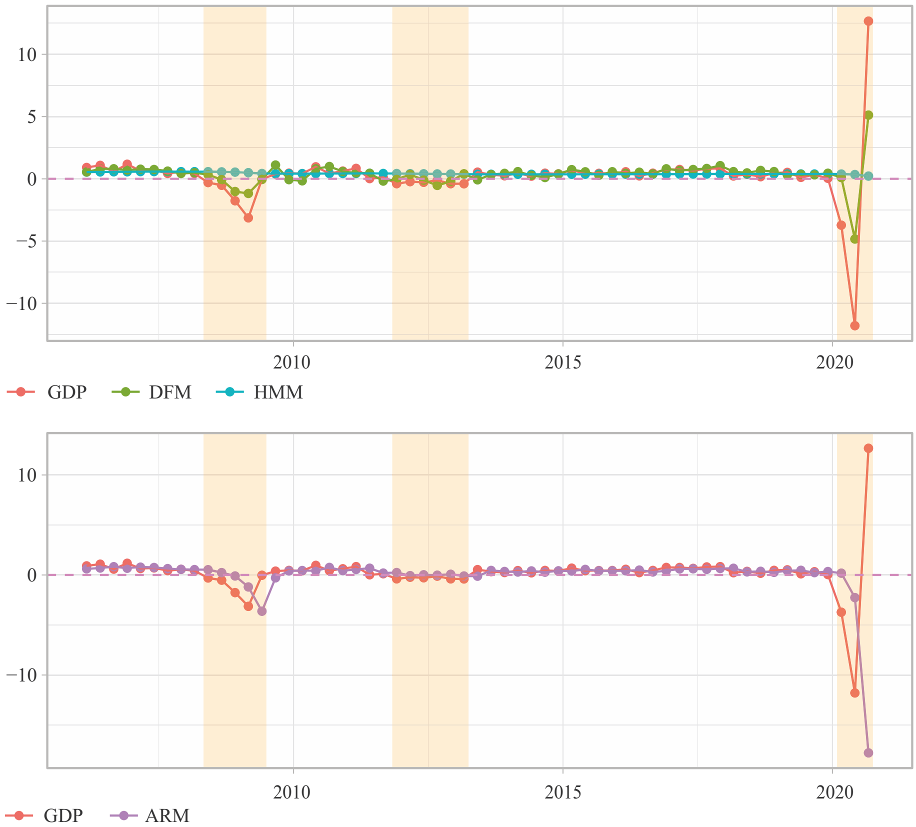
The newly suggested $R^2MSFE(+R)$ metrics is a natural complement to the CSSFED measure. The latter helps identify influential observations that contribute the most to point estimates of the average relative forecast accuracy of the competing models, whereas the former determines the magnitude of the effect these influential observations exert on these estimates.

Very often relative forecasting accuracy is measured using the ratio of root mean squared forecast errors (RMSFE). It is straightforward to modify Equation (2) in order to define the *recursive relative root MSFE* (based on rearranged observations) or $R^3MSFE\ (+R)$:

$$R^3MSFE(+R)_{1/2}^{[j_1,\dots,j_i,\dots,j_P]} = \frac{\sqrt{\sum_{j=j_1}^{j_P} e_{1,j}^2}}{\sqrt{\sum_{j=j_1}^{j_P} e_{2,j}^2}} - 1 \quad (3),$$

where observations are rearranged in the same order as in Equation (2).

Figure 4
GDP growth
(actual values and forecasts)



6 RESULTS

In this section, we compare the forecasting performance of the DFM specified in Section 3 with that of a historical mean model (HMM). Our choice of the HMM as a benchmark model over the more widely used AR(2) model (ARM) in similar exercises (Carriero et al. (2015)) is motivated by the following considerations. First, the HMM forecasts are easy to compute and they are equally easy to communicate to the general public. Second, the HMM offers forecasting performance that is rather similar to that of the ARM for most of the observations during expansions and is typically better than the ARM during sharp upswings in the GDP growth during recovery periods when, by construction, the ARM forecasts underestimate outturns, as these are based on the strongly negative values from the preceding recession period. Third, the ARM model does not pass the COVID-19 test in the sense that, when estimated on the sample ending in the second quarter of 2020, the model coefficient values underwent a drastic change so as to imply an explosive model dynamics. As a result, the AR(2) model predicts strongly negative growth of −17.8 percentage points in the third quarter of 2020 as opposed to the actual GDP growth outturn of 12.7 percentage points. In contrast, the HMM-based forecast for the third quarter of 2020 is 0.2%. Forecasts based on all three models and the corresponding GDP growth outturns are provided in Figure 4.

In any case, it is not that important which benchmark model is chosen for evaluation of relative forecast accuracy during recessions, as the previous research suggests that if a sophisticated multiple-indicator model has any edge over benchmark models in terms of its OOS predictive ability at all than it is during recessions (Chauvet and Potter (2013); Siliverstovs (2020); Siliverstovs and Wochner (2020)). In real-life decision-making, one also rarely relies on forecasts produced by univariate benchmark models, as often the soft information like rumours or stories helps sensing economic distress in the air; moreover, the hard information from high-frequency non-traditional and alternative indicators can be readily incorporated into the forecast generation. The choice of a more robust benchmark model gains in prominence during economic expansions when forecasts from benchmark models are hard to beat (Chauvet and Potter (2013)).

It is also worthwhile mentioning that the DFM coefficient estimates underwent some changes when the observations for the second and third quarters of 2020 were included in the estimation sample. Even in the presence of these changes, the DFM model was still able to produce a forecast for the third quarter of 2020 that was much closer to the outturn than the ARM forecast, as discussed above. The results reported in this section are based on the recursively estimated coefficients of the DFM. We, however, verify the robustness of this approach by generating forecasts from the DFM using coefficient estimates frozen at the pre-COVID period values. The latter approach was also taken in the real-time nowcasting project implemented at the Federal Reserve Bank of New York (Giannone et al. (2017)) based on a dynamic factor model described in Bok et al. (2018).

The point estimates of forecast accuracy are reported in Tables 2 and 3 for the full OOS and the pre-COVID OOS ending in the fourth quarter of 2019. Upon comparing these two tables, one can make the following conclusions on the effect of extending the OOS with the COVID period. First, the values of MSFE substantially increased for all models for the full period as well as for the recessions. For example, for the DFM model we observe a more than 10-fold increase in MSFE for the full period and an about 15-fold increase during recessions. Second, the relative improvement of the DFM over the HMM did not change that much. The $rMSFE_{DFM/HMM}$ reported for the full sample practically remained the same, indicating a reduction in MSFE of about 60%, and it slightly dropped from -0.731 observed for the recessionary periods in the pre-COVID period to -0.631 for all the recessions, including the COVID period. Third, the HMM delivers the highest forecast accuracy during expansions. Both the DFM and ARM report MSFEs about 33% and 22% higher than the HMM. Fourth, the inclusion of the COVID period in the OOS changed the relative ranking of the ARM and HMM models during recessionary sub-periods. In the pre-COVID period, the ARM reported MSFE about 9% lower than the HMM, while for the full set of recessions the ARM model fared much worse than the HMM, reporting an increase in MSFE by about 200%. Based on these observations, the HMM appears a better candidate for the role of a benchmark model. Compared to the ARM, it yields lower MSFEs in expansions (in fact, the **lowest** MSFEs in expansions) and is less affected by the extreme swings in the GDP growth rate observed during the COVID pandemic period.

Table 2
Forecast accuracy
(Q1 2006–Q3 2020)

Model	Full sample		Expansions		Recessions	
	MSFE	rMSFE	MSFE	rMSFE	MSFE	rMSFE
HMM	5.859	0.000	0.066	0.000	24.482	0.000
DFM	2.209	–0.623	0.088	0.333	9.030	–0.631
ARM	17.944	2.062	0.080	0.221	75.365	2.078

Note: rMSFE is computed using the HMM model as the benchmark.

Table 3
Forecast accuracy
(Q1 2006–Q4 2019)

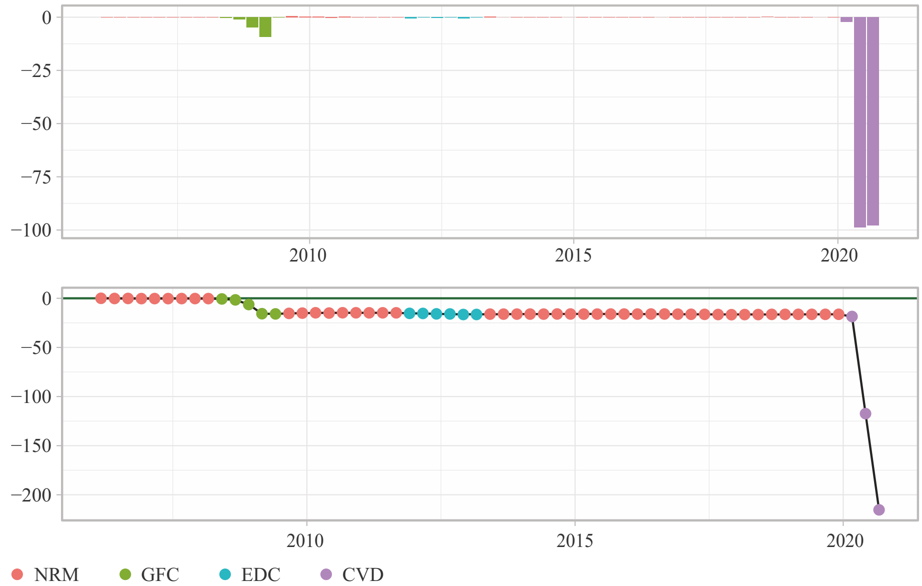
Model	Full sample		Expansions		Recessions	
	MSFE	rMSFE	MSFE	rMSFE	MSFE	rMSFE
HMM	0.474	0.000	0.066	0.000	2.146	0.000
DFM	0.184	–0.613	0.088	0.333	0.577	–0.731
ARM	0.449	–0.053	0.080	0.221	1.959	–0.087

Note: rMSFE is computed using the HMM model as the benchmark.

The nominal and relative forecasting performance summarised in Tables 2 and 3 is based on the averages reported for the full period as well as for its expansionary and recessionary sub-periods. An additional insight into the role of individual observations can be gained from the recursive forecast evaluation metrics discussed in Section 5.

The SFED computed for each observation in the OOS is reported in the upper panel of Figure 5 for the DFM and HMM. The largest differences in squared forecast errors are observed in the fourth quarter of 2008 and the first quarter of 2009 during the GFC period and in the second and third quarters of 2020 during the COVID pandemic. The corresponding $CSSFED_{DFM/HMM}$ is shown in the lower panel of the figure. The displayed CSSFED is characterised by stepwise jumps in these quarters and it conspicuously displays how the difference in forecasting accuracy cumulated over time. Observe that at this scale improvements in the forecast accuracy of the DFM as compared to the HMM during the European Debt Crisis are barely noticeable.

Figure 5
$SFED_{DFM/HMM}$ and $CSSFED_{DFM/HMM}$

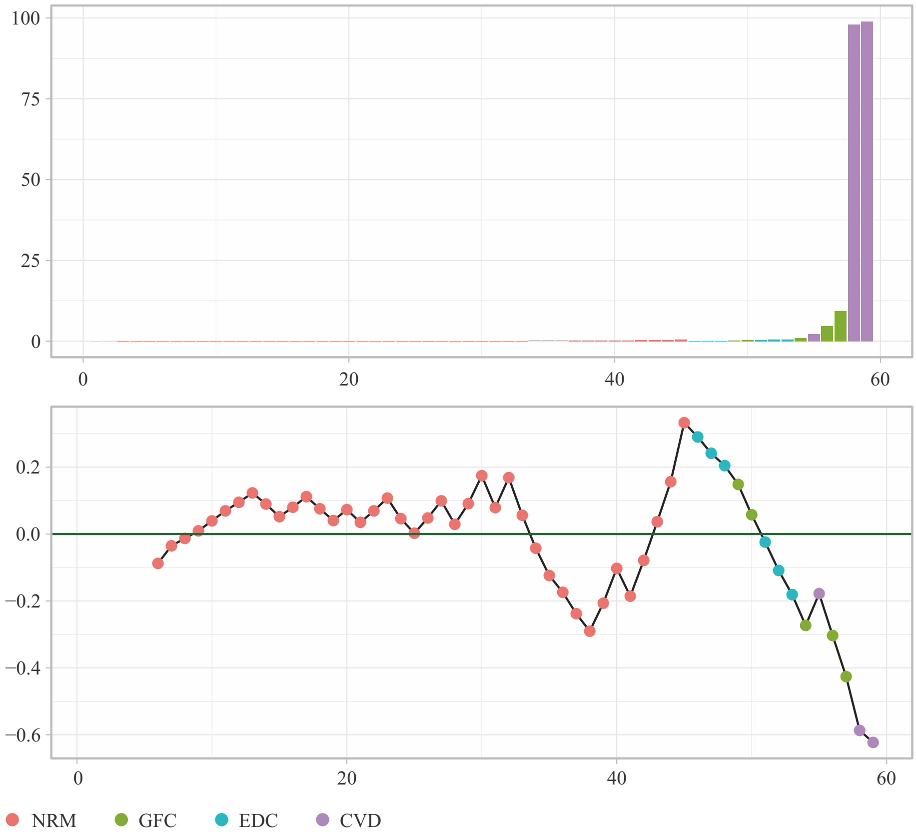


$R^2MSFE(+R)$ suggested in this paper is displayed in the lower panel of Figure 6. $R^2MSFE(+R)$ is based on the observations rearranged according to the phase of the business cycle and the absolute value of SFED and is displayed in the upper panel of the figure. The observations in the expansionary phase (NRM) of the business cycle are shown in red. The last red dot in the $R^2MSFE(+R)$ sequence indicates the value $rMSFE_{DFM/HMM} = 0.333$ reported in Table 2 for expansions. The displayed $R^2MSFE(+R)$ sequence makes it clearly visible how the superiority of the benchmark model in terms of the forecasting ability observed during expansionary periods is gradually eroded as one adds observations from the recessionary periods. The last dot in the $R^2MSFE(+R)$ sequence corresponds to $rMSFE_{DFM/HMM} = -0.623$ reported in Table 2 for the full sample.

It is interesting to observe that all but one recessionary observations contributed to lowering of the relative MSFE. The marginal effect of this one observation in the first quarter of 2020 (i.e. the first quarter of the COVID pandemic) that seemingly counter-intuitively moves the relative MSFE in the opposite direction deserves an explanation, especially given that the DFM produced a lower squared forecast error at this data point, i.e. $SFED_{DFM/HMM,2020Q1} = 14.7 - 16.9 = -2.2 < 0$.

Figure 6

Rearranged Absolute $SFED_{DFM/HMM}$ and $R^2MSFE(+R)_{DFM/HMM}$



Note: The recursive relative MSFE is based on the absolute SFEDs arranged in the ascending order within each business cycle phase (expansions and recessions).

The marginal change in $rMSFE_{1/2}$ caused by an additional data point can be decomposed as follows:

$$\begin{aligned} rMSFE_{1/2}^{P+1} - rMSFE_{1/2}^{P} &= \left[\frac{\sum_{t=1}^{P+1}\hat{e}_{1,t}^2}{\sum_{t=1}^{P+1}\hat{e}_{2,t}^2} - 1\right] - \left[\frac{\sum_{t=1}^{P}\hat{e}_{1,t}^2}{\sum_{t=1}^{P}\hat{e}_{2,t}^2} - 1\right] = \\ &= \left[\frac{\sum_{t=1}^{P}\hat{e}_{1,t}^2}{\sum_{t=1}^{P}\hat{e}_{2,t}^2 + \hat{e}_{2,P+1}^2} + \frac{\hat{e}_{1,P+1}^2}{\sum_{t=1}^{P+1}\hat{e}_{2,t}^2}\right] - \left[\frac{\sum_{t=1}^{P}\hat{e}_{1,t}^2}{\sum_{t=1}^{P}\hat{e}_{2,t}^2}\right] = \\ &= \underbrace{\left[\frac{\sum_{t=1}^{P}\hat{e}_{1,t}^2}{\sum_{t=1}^{P}\hat{e}_{2,t}^2 + \hat{e}_{2,P+1}^2} - \frac{\sum_{t=1}^{P}\hat{e}_{1,t}^2}{\sum_{t=1}^{P}\hat{e}_{2,t}^2}\right]}_{<0} + \underbrace{\frac{\hat{e}_{1,P+1}^2}{\sum_{t=1}^{P+1}\hat{e}_{2,t}^2}}_{>0}. \end{aligned}$$

As can be seen, the total effect depends on the two terms acting in different directions, such that the sign of $rMSFE_{1/2}^{P+1} - rMSFE_{1/2}^{P}$ depends on which partial effect prevails. The sign of the change in relative MSFE does not always conform with the sign of $SFED_{1/2,P+1}$ as it is always the case with changes in CSSFED (see Figure 5).

7 COVID-19 AND STABILITY OF COEFFICIENT ESTIMATES

In the introduction, we mentioned that the large swings in the GDP growth rate observed during the COVID-19 period could not be easily accommodated by econometric models; therefore, in order to avoid deterioration in their forecast accuracy during these turbulent times, some *ad hoc* amendments are necessary to correct for these potentially detrimental consequences (Lenza and Primiceri (2020)). In this section, we verify whether such considerations apply to the DFM used in this study.

The rest of this section is organised as follows. First, we report model coefficient estimates using an expanding window that rolls over the COVID-19 period. We focus particularly on the estimates of the autoregressive parameter ϕ in the equation for the common factor f_t and the factor loading coefficient β in the equation of the variable of our main interest (GDP) and the loading coefficients λ^i to each of the auxiliary monthly indicators $i \in \{IIP, EXI, EXJ, IOR, ESI\}$. Second, we freeze the coefficient values at their estimates when targeting GDP growth in the fourth quarter of 2019 (i.e. the last quarter in the pre-COVID period) and use these coefficient values to generate forecasts of the GDP growth rate in the COVID period. This approach of freezing coefficient values at their pre-COVID estimates in response to the extraordinary movements in the US GDP data in the second and third quarters of 2020 was, for example, adopted in the nowcasting model of the Federal Reserve Bank of New York (Bok et al. (2018)) as well as in the resuscitated mixed-frequency Bayesian VAR model (Schorfheide and Song (2020)). Third, by comparing the forecast accuracy from the model with recursively estimated coefficients and the model with frozen coefficients we can determine which approach resulted in better forecast accuracy.

The DFM coefficient values that were recursively estimated are reported in Table 4. As anticipated, inclusion of observations of the auxiliary monthly variables over the months of the second quarter of 2020 in the estimation period results in noticeable changes in the values of the estimated coefficients. The autoregressive coefficient measuring factor persistence decreased, whereas all the loading coefficients (β and λ^i's) substantially gained in their value. This observation is in line with the argument that the large and synchronous swings in the auxiliary indicators improved the identification of the latent factor. Next, we verify whether these changes in the coefficient estimates result in more accurate forecasts during the COVID period compared with the forecasts generated from the model with frozen estimates.

Table 4
Recursively estimated coefficients

Target quarter		φ	β	λ^{IIP}	λ^{EXI}	λ^{EXJ}	λ^{IOR}	λ^{ESI}
Q4 2019	coefficient	0.8192	0.1496	0.2547	0.2589	0.2020	0.2500	0.0660
	SE	0.0397	0.0195	0.0370	0.0374	0.0370	0.0369	0.0057
Q1 2020	coefficient	0.8092	0.1595	0.2727	0.2815	0.2205	0.2683	0.0700
	SE	0.0440	0.0214	0.0412	0.0417	0.0403	0.0406	0.0067
Q2 2020	coefficient	0.1570	0.5337	0.8439	0.7910	0.7013	0.7709	0.2211
	SE	0.0629	0.0664	0.0447	0.0434	0.0472	0.0469	0.0110
Q3 2020	coefficient	0.2834	0.4933	0.8210	0.7727	0.6962	0.7590	0.2260
	SE	0.0581	0.0392	0.0435	0.0418	0.0451	0.0452	0.0113

Note: Model coefficients were estimated using an expanding window of observations starting in the first quarter of 1995 and ending in the target quarter. At each forecast origin, the value of GDP growth for the previous quarter was included in the information set.

The DFM forecasts generated using the recursively estimated coefficients and the coefficients frozen at their pre-COVID estimates are shown in Table 5 together with the actual GDP growth figures reported in the latest vintage at our disposal. When forecasting the GDP growth in the fourth quarter of 2019, the forecasts are identical since these are based on the same coefficient values. For the first quarter of 2020, the forecasts are so close that these cannot be distinguished at this level of rounding. The differences become noticeable when forecasting the GDP growth in the second and third quarters of 2020. The forecasts based on frozen coefficients are more conservative, whereas the forecasts from the recursively estimated model turned out to be closer to the outturn values. Thus, the DFM copes successfully with the structural changes in the values of its coefficient estimates caused by extreme swings in the data during the COVID period and produces forecasts that are superior to forecasts based on frozen parameter estimates.

Table 5
GDP outturns and forecasts in the last four quarters of the OOS period

Target quarter	DFM forecasts using		GDP growth
	recursive estimates	frozen estimates	outturn
Q4 2019	0.4	0.4	0.0
Q1 2020	0.1	0.1	–3.7
Q2 2020	–4.8	–3.7	–11.8
Q3 2020	5.1	2.7	12.7

Note: DFM forecasts obtained using recursive estimates is a subset of the forecasts analysed in the main text. DFM forecasts obtained using frozen estimates are forecasts generated by the model with parameters frozen at their pre-COVID values (namely, at those of the fourth quarter of 2019).

8 CONCLUSIONS

In this paper, we presented the results of forecasting the euro area GDP growth over the period from the first quarter of 2006 to the third quarter of 2020, paying a special attention to the models' forecasting performance during the COVID-19 pandemic. Extremely large swings in the GDP growth rate observed in the second and third quarters of 2020 imply that the forecast errors of the econometric forecasting models for these quarters are also highly likely to be extraordinarily large. Undoubtedly, these large forecast errors exert very large leverage on the forecast accuracy metrics based on averages of squared forecast errors and their differentials. In order to avoid bias when ranking models based on their forecast accuracy, it is better to rely on recursive forecast evaluation metrics that dissect a model's forecasting performance observation by observation. The recursive metrics, referred as the cumulative sum of squared forecast error difference (CSSFED) suggested in Welch and Goyal (2008), allows to detect influential observations that are usually pivotal in ranking models based on their average forecasting performance.

We introduce another recursive forecast evaluation metrics in this paper, $R^2MSFE(+R)$, that is complementary to the CSSFED metrics. Its main purpose is to directly track changes in relative forecast accuracy caused by such influential observations. When used together, these two metrics enable a better understanding of the sources and causes of one model's forecasting supremacy over its competitors as well as the timing when this happens. It also allows to gauge the robustness of the models' relative ranking to omission of one or several influential observations.

We illustrate the usefulness of such recursive metrics as CSSFED and $R^2MSFE(+R)$ when forecasting the euro area GDP growth during the COVID-19 pandemic as well as the two previous recessionary periods covered by our data sample. It turned out that the DFM used for generating forecasts successfully coped with the data challenges posed by the COVID-19 pandemic. The forecasts based on the recursively estimated coefficients proved out to be much closer to the outturns of GDP growth in the second and third quarters of 2020 than the forecasts based on the coefficients frozen at their pre-COVID period values (see Schorfheide and Song (2020) for their experience when forecasting the US economic growth during the COVID-19 pandemic).

BIBLIOGRAPHY

ANTOLIN-DIAZ, Juan, DRECHSEL, Thomas, PETRELLA, Ivan (2017). Tracking the Slowdown in Long-Run GDP Growth. *The Review of Economics and Statistics*, vol. 99, issue 2, pp. 343–56.

ANTOLIN-DIAZ, Juan, DRECHSEL, Thomas, PETRELLA, Ivan (2020) Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks, and New Data. [Paper presented at the World Congress of the Econometric Society, 2020].

BANBURA, Marta, BOBEICA Elena (2020). *Does the Phillips Curve Help to Forecast Euro Area Inflation?* Working Paper Series. European Central Bank. No. 2471, September 2020, 55 p.

BAUMEISTER, Christiane, GUERIN, Pierre (2020). *A Comparison of Monthly Global Indicators for Forecasting Growth*. NBER Working Paper, No. 28014, October 2020, 36 p.

BOK, Brandyn, CARATELLI, Daniele, GIANNONE, Domenico, SBORDONE, Argia M., TAMBALOTTI, Andrea (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review of Economics*, vol. 10, issue 1, pp. 615–643.

CAMACHO, Maximo, PEREZ-QUIROS, Gabriel (2010). Introducing the EURO-STING: Short-Term Indicator of Euro Area Growth. *Journal of Applied Econometrics*, vol. 25, issue 4, pp. 663–694.

CARRIERO, Andrea, CLARK, Todd E., MARCELLINO Massimiliano (2015). Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. *Journal of the Royal Statistical Society*. Series A (Statistics in Society), vol. 178, issue 4, October 2015, pp. 837–862. Available: *https://doi.org/10.1111/rssa.12092*.

CHAUVET, Marcelle, POTTER, Simon (2013). Forecasting Output. ***In:*** Elliott, Graham, Granger, Clive, Timmermann, Allan (ed.), *Handbook of Forecasting*, vol. 2, chapter 3, pp. 141–194.

CIMADOMO, Jacopo, GIANNONE, Domenico, LENZA, Michele, SOKOL, Andrej, MONTI, Francesca (2020). *Nowcasting with large Bayesian vector autoregressions.* Working Paper Series No. 2453, August 2020. European Central Bank. 30 p.

DELLE MONACHE, Davide, DE POLIS, Andrea, PETRELLA, Ivan (2020). *Modelling and Forecasting Macroeconomic Downside Risk*. EMF Research Papers No. 34. Economic Modelling Forecasting Group.

GEWEKE, John, AMISANO, Gianni (2010). Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting*, vol. 26, issue 2, pp. 216–230. Available: *https://doi.org/10.1016/j.ijforecast.2009.10.007*.

GIANNONE, Domenico, SBORDONE, Argia, TAMBALOTTI, Andrea (2017). *Hey, Economist! How Do You Forecast the Present?* Federal Reserve Bank of New York Liberty Street Economics (blog). Available: *http://libertystreeteconomics.newyorkfed.org/2017/06/hey-economist-how-do-you-forecast-the-present.html*.

LENZA, Michele, PRIMICERI Giorgio E. (2020). *How to estimate a VAR after March 2020*. Working Paper Series No. 2461, August 2020. European Central Bank. 20 p.

MARCELLINO, Massimiliano, SCHUMACHER, Christian (2010). Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, August 2010, vol. 72, issue 4, pp. 518–550.

MARIANO, Roberto S., MURASAWA, Yasutomo (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, vol. 18, issue 4, pp. 427–443.

PEREZ-QUIROS, Gabriel, ROTS, Eyno, LEIVA-LEON, Danilo (2020). *Real-Time Weakness of the Global Economy: A First Assessment of the Coronavirus Crisis*. Working Paper Series No. 2381, March 2020. European Central Bank. 60 p.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna : R Foundation for Statistical Computing. Available: *http://www.R-project.org/*.

SCHORFHEIDE, Frank, SONG Dongho (2020). *Real-Time Forecasting with a (Standard) Mixed-Frequency VAR During a Pandemic*. Working Papers No. WP20–26, July 2020. Federal Reserve Bank of Philadelphia. 14, A-3 p. Available: *https://doi.org/10.21799/frbp.wp.2020.26*.

SILIVERSTOVS, Boriss (2017). Dissecting models' forecasting performance. *Economic Modelling*, vol. 67, issue C, pp. 294–299.

SILIVERSTOVS, Boriss (2020a). Assessing Nowcast Accuracy of US GDP Growth in Real Time: The Role of Booms and Busts. *Empirical Economics*, Springer, vol. 58, issue 1, No. 2, January 2020, pp. 7–27.

SILIVERSTOVS, Boriss, WOCHNER, Daniel S. (2020b) State-Dependent Evaluation of Predictive Ability. *Journal of Forecasting* (forthcoming).

WELCH, Ivo, GOYAL Amit (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, vol. 21, issue 4, July 2008, pp. 1455–1508.