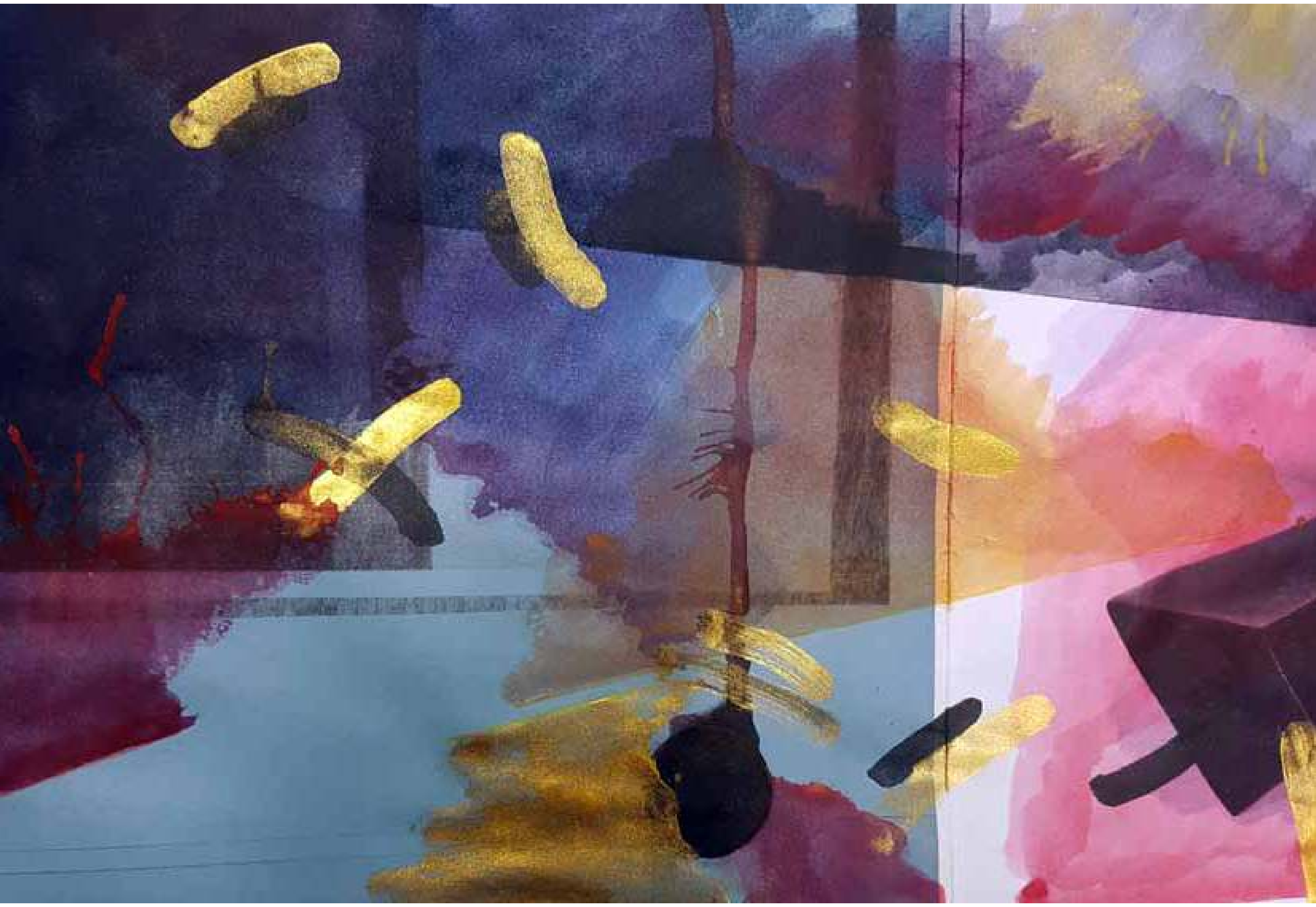
ISBN-978-9934-578-05-2

BORISS SILIVERSTOVS

WORKING PAPER
1 / 2019

ASSESSING NOWCAST ACCURACY OF US GDP GROWTH IN REAL TIME: THE ROLE OF BOOMS AND BUSTS



CONTENT

ABBREVIATIONS

ALS – average logarithmic score
ALSD – average logarithmic score difference
AR – autoregression
CCM – Carriero, Clark, Marcellino
CDF-EW – combination density forecasts – equal weighting
CDF-RW – combination density forecasts – recursive weighting
CSSFED – cumulated squared sum of forecast error difference
CSLSD – cumulated sum of logarithmic score difference
GDP – gross domestic product
FO – forecast origin
NBER – National Bureau of Economic Research
OLS – ordinary leat squares
RMSFE – root mean squared forecast error
RRMSFE – relative root mean squared forecast error
RW – random walk
SIM – single-indicator model
SV – stochastic volatility
UNI-AR – univariate autoregression
US – United States of America
vs – versus

ABSTRACT

In this paper we reassess the forecasting performance of the Bayesian mixed-frequency model suggested in Carriero et al. (2015) in terms of point and density forecasts of the GDP growth rate using US macroeconomic data. Following Chauvet and Potter (2013), we evaluate the forecasting accuracy of the model relative to a univariate AR(2) model separately for expansions and recessions, as defined by the NBER business cycle chronology, rather than relying on a comparison of forecast accuracy over the whole forecast sample spanning from the first quarter of 1985 to the third quarter of 2011. We find that most of the evidence favouring the more sophisticated model over the simple benchmark model is due to relatively few observations during recessions, especially those during the Great Recession. In contrast, during expansions the gains in forecasting accuracy over the benchmark model are at best very modest. This implies that the relative forecasting performance of the models varies with business cycle phases. Ignoring this fact results in a distorted picture: the relative performance of the more sophisticated model in comparison with the naive benchmark model tends to be overstated during expansions and understated during recessions.

Keywords: nowcasting, mixed-frequency data, real-time data, business cycle

JEL code: C22, C53

The paper benefited from comments made by participants of a seminar held at the Department of Economics and Finance, the Tallinn University of Technology (Estonia), by participants of the 18th IWH-CIREQ-GW Macroeconometric Workshop: Mixed Frequency Data in Macroeconomics and Finance (Halle, Germany), the 22nd Annual International Conference on Macroeconomic Analysis and International Finance (Rethymno, Greece), the 1st Baltic Economic Conference (Vilnius, Lithuania) as well as the 1st Vienna Workshop on Economic Forecasting 2018 (Austria). All computations were performed in R and RcppArmadillo. Any remaining errors are the sole responsibility of the author – Chief Econometrician of the Monetary Policy Department of Latvijas Banka and a Research Fellow at the KOF Swiss Economic Institute, ETH Zurich, Switzerland.

Boriss Siliverstovs: Latvijas Banka, K. Valdemāra iela 2A, Riga, LV-1050, Latvia; e-mail: Boriss.Siliverstovs@bank.lv, KOF Swiss Economic Institute at ETH Zurich, 8092 Zurich, Switzerland.

1. INTRODUCTION

In a recent contribution to the *Handbook of Forecasting* Chauvet and Potter (2013) provided a comprehensive review of the forecasting performance of several state-of-the-art econometric models using real-time data vintages. These models include the univariate linear autoregressive model of order two, AR(2), univariate non-linear models that take into account Cumulative Depth of Recession (CDR) and Markov-switching dynamics, the Dynamic Stochastic General Equilibrium (DSGE) model, Bayesian and non-Bayesian Vector Auto Regressive (BVAR and VAR) models as well as the Dynamic Factor model with Markov Switching (AR-DFMS). In addition to forecasts from these econometric models, survey-based Blue Chip forecasts were also included in the analysis.

Focusing on the quarterly US GDP growth rate, Chauvet and Potter (2013) conducted an assessment of the forecasting accuracy of these models not only for the full sample (from the first quarter of 1992 to the fourth quarter of 2010) but also for sub-samples of expansions and recessions, as defined by the NBER business cycle chronology. Their main finding is that forecasting performance of the state-of-the-art models varies with the business cycle phases. Typically, the absolute size of forecast errors is larger during recessions than expansions. More intriguingly, during expansions all sophisticated models in question as well as professional forecasters produced average forecast accuracy that at best is comparable to that from the benchmark AR(2) model.[1]

Despite the clearly stated outcome, the suggestion of Chauvet and Potter (2013) to report absolute and relative forecasting performance of models separately for recessions and expansions has so far not received much attention in the macroeconomic forecasting literature. Recent studies persist in reporting results of forecasting competitions for the whole forecast evaluation period (e.g. see Schorfheide and Song (2015), Kim and Swanson (2018)) or, at best, for the sub-periods that either end or even start just before the Great Recession (e.g. see Carriero et al. (2015), Foroni et al. (2015), Giannone et al. (2016)). In doing so, these studies are likely to conceal differences in forecasting performance across recessions and expansions of competing models, potentially leading to the erroneous conclusions regarding the ranking of these models based on their relative predictive ability.

In this paper we investigate whether the conclusions reached in Chauvet and Potter (2013) for forecasting models involving variables sampled at a single (quarterly) frequency can be generalised to models dealing with economic variables sampled at heterogeneous frequencies. Since the seminal work of Ghysels et al. (2004) and Ghysels et al. (2007), mixed-frequency models have gained enourmous popularity among the forecasting community, with many different modifications proposed to the original model specifications (Siliverstovs (2017), Carriero et al. (2015), Foroni et al. (2015), Schorfheide and Song (2015), Guérin and Marcellino (2013), Marcellino and Schumacher (2010), *inter alia*); see Foroni and Marcellino (2013) for an overview. However, to the best of our knowledge the question of the comparative predictive ability of mixed-frequency models during economic booms and busts has not yet been addressed in a systematic way. In our study we intend to close this gap in the literature by providing detailed empirical evidence on this topic.

[1] Similar to Chauvet and Potter (2013), the AR(2) model was chosen as the benchmark model in studies by Carriero et al. (2015), Edge et al. (2010), Siliverstovs (2017) and many others.

For this purpose, we utilise the Bayesian mixed-frequency model with stochastic volatility proposed in Carriero et al. (2015) for forecasting US GDP growth using the 12 most closely monitored monthly economic/financial indicators. Our choice of this study for emphasising the importance of verifying asymmetries in the predictive ability of competing models across the business cycle phases is not purely incidental. First, the econometric model of Carriero et al. (2015) combines several recent advances in time series econometrics that makes it suitable for routine forecasting in institutions such as central banks, for example. These include handling of mixed-frequency data sets of moderate size, straightforward and fast parameter estimation using Bayesian methods, out-of-sample model evaluation both in terms of point and density forecast accuracy and, last but not least, incorporation of stochastic volatility of the error term. The last feature has been shown to be instrumental in improving density forecasts of US macroeconomic time series compared to models with homoscedastic innovations (Clark (2011)).

Second, the data set of Carriero et al. (2015) comprises historical real-time data vintages for each month from January 1985 until October 2011, thus avoiding the caveats of using pseudo-real time data sets for model estimation and evaluation (Croushore and Stark (2003; 2001)). More importantly, in line with common practice Carriero et al. (2015) evaluate the forecasting performance of their model over the whole forecast sample from the first quarter of 1985 until the third quarter of 2011 as well as for the sub-sample from the first quarter of 1985 until the first quarter of 2008 that ends before the Great Recession. Using the latter sub-sample, the authors exclude one out of three NBER-identified recession periods, while in the former sample all three recessions are present. All in all, the results reported in Carriero et al. (2015) serve as a well-documented, credible benchmark against which we can compare our results.

Our main findings can be summarised as follows. First, similarly to Chauvet and Potter (2013) we document asymmetric forecasting ability during expansions and recessions. But in our case this asymmetry gradually vanishes as the forecasting horizon shortens and more information regarding the reference quarter accrues from monthly indicators. Remarkably, this conclusion holds not only for point but also for density forecasts.

Second, even though differences in forecasting accuracy during expansions/recessions eventually disappear for models based on the most complete information sets, the differences in forecasting performance relative to the benchmark model remain. Consistent with the results of Chauvet and Potter (2013), we also find that during expansions the benchmark univariate autoregressive model produces forecasting accuracy that is comparable to that of more sophisticated multivariate models. It is only for relatively few observations during recessions that these models are able to bring about substantial gains in forecast accuracy relative to the benchmark model.

This recorded asymmetry in the relative forecasting performance across the business cycle phases has strong implications for the message delivered by those studies that ignore it. In doing so, these studies tend to severely overstate the predictive ability of their preferred models over that of naive benchmark models during expansions and, consequently, to understate it during recessions. Hence, the biased assessment of model forecasting accuracy is delivered to business analysts, policy-makers or any other parties interested in their forecasts.

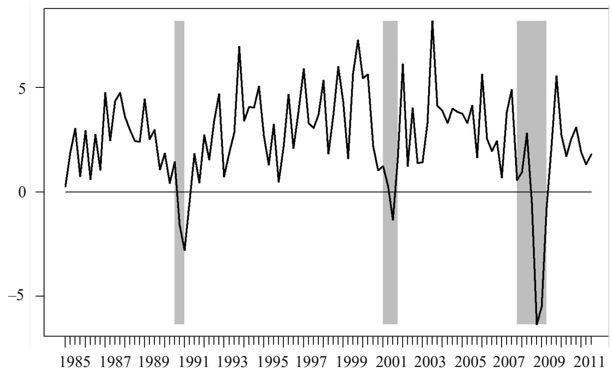
The rest of the paper is organised as follows. The next section contains a description of the data set. Section 3 describes the econometric model used in the paper. In Section 4 we present evaluation metrics used for assessment of forecasting performance of the forecasting models. The results are presented in Section 5. The results reported in the main text are complemented in the extensive Appendix, where we verify the robustness of our conclusions by assessing forecasting performance of alternative benchmark models, SIMs and their combinations. The final section summarises our findings and outlines possible extensions of our work.

2. DATA

The data used in this paper are the real-time data vintages collected by Carriero et al. (2015). These vintages comprise quarterly GDP data as well as 12 monthly indicators that are closely monitored for assessing economic outlook in the US. The data set is organised to reflect historical availability of both quarterly GDP data and monthly indicators for every month from January 1985 until October 2011. These monthly data vintages are utilised in order to produce US GDP growth forecasts for each quarter from the first quarter of 1985 until the third quarter of 2011.

The US GDP data for quarter $t-1$ are published in three subsequent releases (initial, second, and final) at the end of each month of the following quarter t. In line with Carriero et al. (2015), we evaluate the forecasting accuracy of the econometric models using the second release of the GDP data (see Figure 1). The shaded areas correspond to the three recession periods according to the NBER business cycle chronology: from the third quarter of 1990 until the first quarter of 1991 (three quarters), from the first quarter of 2001 until the fourth quarter of 2001 (four quarters), and from the fourth quarter of 2007 until the second quarter of 2009 (seven quarters). Thus, in the 107-quarter long forecast evaluation sample there are 14 and 93 quarters identified as recession and expansion periods, respectively.

Figure 1
Quarterly US GDP growth; second release



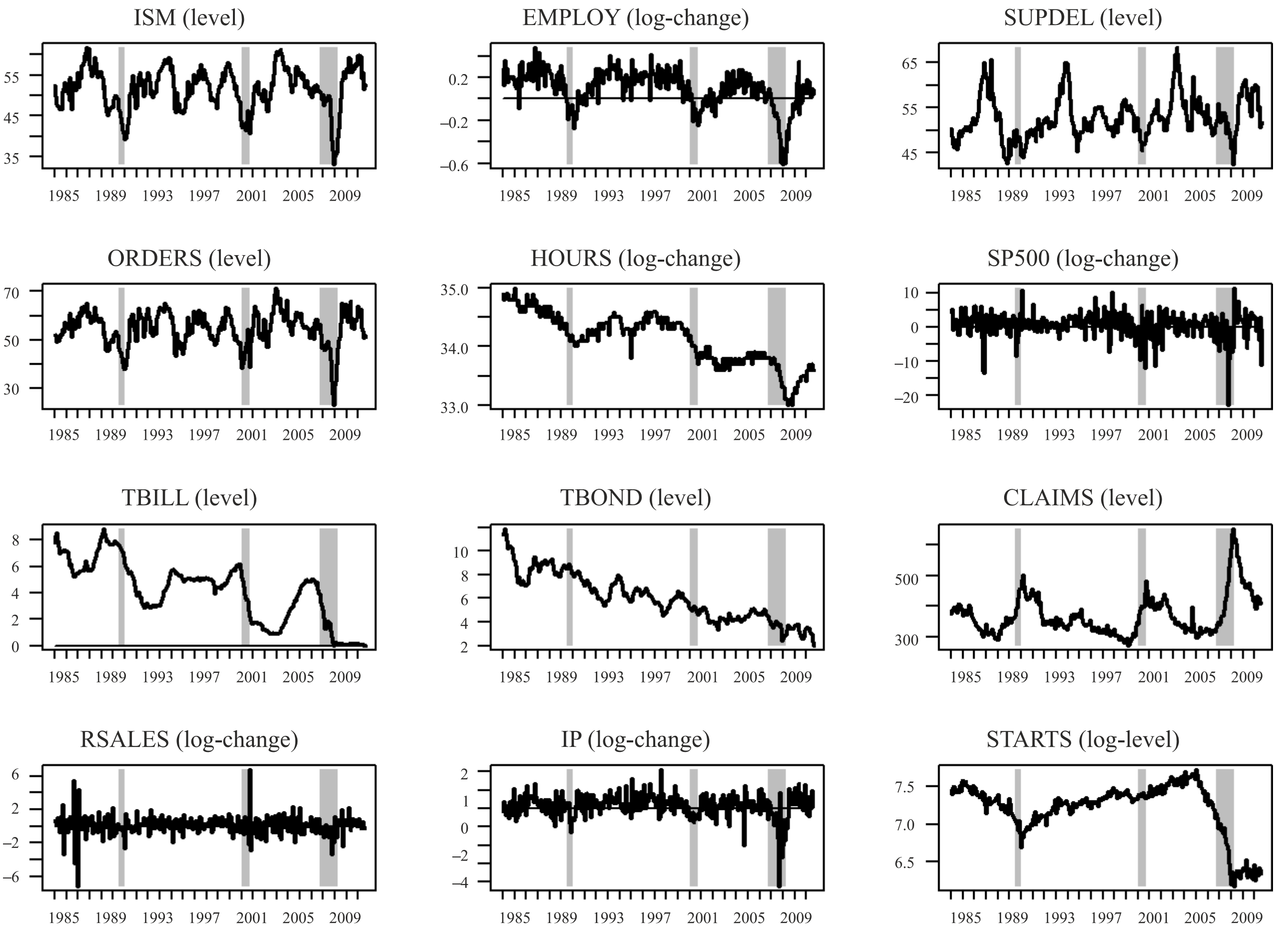
The monthly data include business tendency surveys (ISM, SUPDEL, ORDERS), labour market conditions (EMPLOY, CLAIMS), production (IP, HOURS) as well as consumption (RSALES) measures of the economy, housing (HS) and financial (SP500, TBILL, TBOND) markets. These indicators are characterised by their timing of release, i.e. whether their values for the previous month are released during the first or second week in the current month (see Table 1). The differences in the release timing of the monthly indicators have important implications for the specification of forecasting models at forecasting origins. The time series of monthly indicators from the data vintage (October 2011) along with the recession periods are displayed in Figure 2.

Table 1
Monthly indicators

Name	Description (transformation)	Timing of release[1]
ISM	ISM index (overall) for manufacturing (level)	1st week
EMPLOY	Payroll employment (log-change)	1st week
SUPDEL	ISM index for supplier delivery times (level)	1st week
ORDERS	ISM index for orders (level)	1st week
HOURS	Average weekly hours of production workers (log-change)	1st week
SP500	S&P 500 index (log-change)	1st week
TBILL	3-month Treasury bill rate (level)	1st week
TBOND	10-year Treasury bond yield (level)	1st week
CLAIMS	New claims for unemployment insurance (level)	2nd week
RSALES	Real retail sales (log-change)	2nd week
IP	Industrial production (log-change)	2nd week
STARTS	Housing starts (log-level)	2nd week

[1] Release of the observation for the previous month either in the first or second week of the current month.

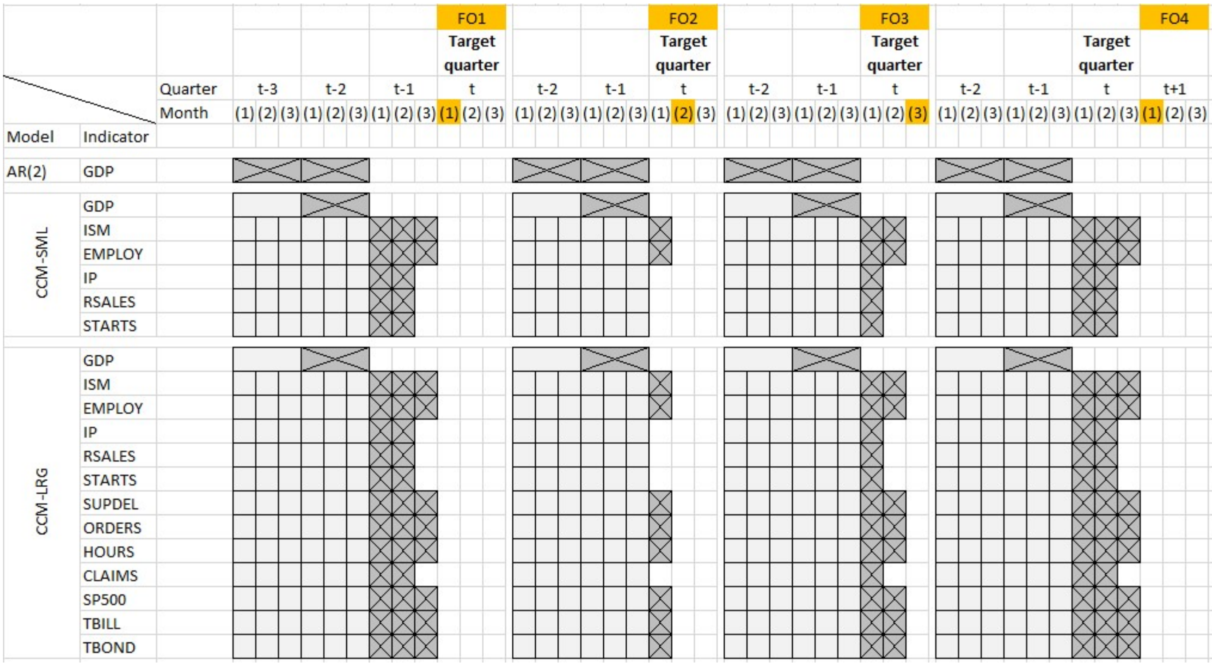
Figure 2
Data: monthly indicators, vintage from October 2011; forecast sample



3. ECONOMETRIC MODEL OF CARRIERO ET AL. (2015)

General information about the forecasting framework is presented in Figure 3. The forecasts of US GDP growth are produced for each quarter in the period from the first quarter of 1985 to the third quarter of 2011 at four consecutive forecast origins (FO1, FO2, FO3, FO4) as follows: FO1, FO2, and FO3 denote the end of the first week of the first, second and third month in quarter t, respectively, and FO4 denotes the end of the first week of the first month in quarter $t + 1$. In line with the rest of the literature, forecasts made at forecast origins FO1–FO3 are referred to as *nowcasts*, i.e. those made for now or the current quarter, and forecasts made at FO4 are labelled as *backcasts*, i.e. those that are backwards-looking or that are made (shortly) after the end of the targeted quarter.

Figure 3
Data release timing and model specifications



The forecasts are made recursively month by month, using an expanding information set. In Figure 3 we use gray-shaded colour in order to indicate data availability at each forecast origin.[2] For example, at FO1 in a given quarter t the values up to the third month of the previous quarter $t - 1$ are available for the variables released during the first week of the month: ISM, SUPDEL, ORDERS, EMPLOY, HOURS, SP500, TBILL, and TBOND. For the remaining four variables (CLAIMS, RSALES, IP, HS) released during the second week of the month, only the values up to the second month of that quarter $t - 1$ are available to the forecaster. At the next forecast origin FO2, the data are extended by one month. This means that for the first and second groups of variables (according to the publication week) observations for the first month of quarter t and for the third month of quarter $t - 1$ have been released, respectively. For each of the remaining forecast origins data availability consequently increases by one month for all monthly variables.

[2] We distinguish between plain and crisscrossed cells both coloured in gray. The description of the crisscrossed cells is provided in Sub-section 3.3 below.

Similarly to the monthly variables, at each forecast origin the availability of GDP data is shown in Figure 3. At FO1 in quarter t, GDP data are available up to quarter $t-2$. At the next forecast origin FO2, the first publication of the official GDP estimate for quarter $t-1$ takes place, followed by the second and third releases for the same quarter $t-1$ at FO3 and FO4, respectively.

The initial estimation sample is from the first quarter of 1970 to the third quarter of 1984 for FO1 and from the first quarter of 1970 to the fourth quarter of 1984 for FO2–FO4, reflecting the publication lag of GDP vintages. Using the available data, a forecast is made for the first quarter of 1985 at each of the four forecast origins. Then the estimation sample is increased by one quarter, the first quarter of 1985, and forecasts for the second quarter of 1985 are made. We proceed in this fashion until forecasts for the last quarter, the third quarter of 2011, in our forecast evaluation sample are made.

3.1 General setup

In this section we describe the econometric model of Carriero et al. (2015). Since we intend to replicate the results of their paper, all model specifications and tuning parameter values are kept the same as in the original paper. The econometric model has the following specification:

$$y_t = X'_{m,t}\beta_m + \varepsilon_{m,t}, \quad (1),$$

$$\varepsilon_{m,t} = \kappa_{m,t}^{0.5}\varepsilon_{m,t}, \varepsilon_{m,t} \sim NIID(0,1) \quad (2),$$

$$\ln\kappa_{m,t} = \ln\kappa_{m,t-1} + \nu_{m,t}, \nu_{m,t} \sim NIID(0,\phi_m) \quad (3).$$

The conditional mean is modelled as a linear function of the explanatory variables collected in vector $X_{m,t}$. The subindex $m = \{FO1, FO2, FO3, FO4\}$ corresponding to one of the forecast origins indicates that data vector $X_{m,t}$ is specific for every forecast origin, reflecting data release timing as discussed in Section 2. In general, the vector $X_{m,t}$ contains an intercept, lags of the dependent variable as well as quarterly values of the original monthly indicators so that the variables on both the left- and right-hand sides of the regression equation are at the quarterly frequency. The conversion of the original monthly indicators to quarterly frequency is achieved by skip-sampling their monthly values. For example, a monthly variable w_t is converted to quarterly frequency by sampling every third observation of w_t in such a way that all the observations pertaining to the first, second and third months in every quarter are collected in three quarterly time series $w_t^{(1)}, w_t^{(2)}, w_t^{(3)}$ where superscripts (i) with $i = 1,2,3$ indicate the first, second or third month in quarter t.

Essentially, the model in equations (1)–(3) can be considered as an extended version of the U-MIDAS model of Foroni et al. (2015), at least in the following two aspects. First, the conditional mean in equation (1) can have an arbitrary number of (skip-sampled) monthly indicators (the maximum is 12 in this application). Observe that the original specification of the U-MIDAS model in Foroni et al. (2015) allows only one monthly indicator at a time as a straightforward generalisation of the non-linear MIDAS models of Ghysels et al. (2004). The second important extension introduces stochastic volatility in the mixed-frequency forecasting model (see equations (2) and (3)). Thus, in the most general form, the conditional variance of the error term $\varepsilon_{m,t}$ is modelled as a time-varying stochastic process. It also can be switched off, resulting in a model with constant volatility of disturbances as in the original (U-)MIDAS model specifications.

Allowing for multiple indicators in the model generally leads to parameter inflation problem, as by means of skip-sampling one high-frequency indicator is converted into p low-frequency regressors where p denotes the frequency mismatch parameter.[3] In such situations OLS estimation of the U-MIDAS model parameters, advocated in Foroni et al. (2015), quickly leads to overfitting. One solution, adopted by Carriero et al. (2015), is to use the Bayesian approach for estimation of the model parameters as well as generation of out-of-sample forecasts. In addition to point forecasts, as a by-product of Bayesian estimation, density forecasts are generated that take into account both the parameter estimation uncertainty and (potentially) time-varying variance of the error term.

3.2 Priors

For the model in equations (1)–(3) we use normal Minnesota-style priors on the coefficient vector β_m, characterised by mean zero and diagonal covariance matrix. The degree of shrinkage is controlled by the three hyperparameters: λ_1 determines the overall rate of shrinkage; λ_2 sets the shrinkage rate of the monthly variables relative to that of the lags of the GDP variable; and λ_3 regulates the shrinkage rate imposed on the longer lags of the regressors. The diagonal entries of the prior covariance matrix for β_m are based on the following:
- for the intercept, $sd_{incpt} = 1000 * \sigma_{y_t}$
- for the lagged dependent variable, $sd_{y_{t-l}} = \lambda_1 / l^{\lambda_3}$
- for the monthly indicators, $sd_{w_{t-l}^{(i)}} = \sigma_{y_t} / \sigma_{w_t^{(i)}} * (\lambda_1 \lambda_2) / l^{\lambda_3}$.

The values σ_{y_t} and $\sigma_{w_t^{(i)}}$ are estimated using regression standard errors of AR(4) models applied to the dependent and explanatory variables, respectively. The hyperparameters are set to $\lambda_1 = \lambda_2 = 0.2$ and $\lambda_3 = 1$, which is common in the literature as mentioned in Carriero et al. (2015, p. 845).

Diffuse priors are set for the variance of the error term for models with constant volatility. For the models with stochastic volatility, the priors on the volatility components are independent of those for the coefficient vector β_m. The prior distribution of ϕ is characterised by mean equal to 0.035 and 5 degrees of freedom. The prior distribution for the initial value of κ_0 is normal, $N(\ln\hat{\kappa}_0, 4)$ where $\hat{\kappa}_0$ is the regression standard error of the AR(4) model fitted to GDP growth.

3.3 Model specifications

Carriero et al. (2015) employ two model specifications, depending on the number of monthly indicators included: five- and 12-indicator models, labelled CCM-SML and CCM-LRG, respectively. The forecasts of these models are compared with those of the benchmark model – the univariate AR(2) model that utilises no external information but its own two lagged values. In Figure 3 the specification of the AR(2) model is denoted by crisscrossed cells that span a whole quarter such that $X_{m,t}$ in

[3] In the case of monthly-quarterly data, p equals three, such that the number of regressors effectively triples.

equation (1) reads
$X_{FO1,t} = (1, y_{t-2}, y_{t-3})'$ and $X_{FO2,t} = X_{FO3,t} = X_{FO4,t} = (1, y_{t-1}, y_{t-2})'$.[4]

Similarly, the model specification of the CCM-SML and CCM-LRG models can be inferred from Figure 3. But observe that in the case of high-frequency economic/financial indicators each crisscrossed cell indicates a separate quarterly time series of skip-sampled monthly values that only contains values in month (i) of each quarter, with $i = 1,2,3$. For example, for the CCM-SML model we have

$X_{FO1,t} = (1, y_{t-2}, ISM_{t-1}^{(3)}, ISM_{t-1}^{(2)}, ISM_{t-1}^{(1)}, EMPLOY_{t-1}^{(3)},$

$EMPLOY_{t-1}^{(2)}, EMPLOY_{t-1}^{(1)}, IP_{t-1}^{(2)}, IP_{t-1}^{(1)}, RSALES_{t-1}^{(2)},$

$RSALES_{t-1}^{(1)}, STARTS_{t-1}^{(2)}, STARTS_{t-1}^{(1)})'$, with in total 14 regressors. At the next forecasting origin the vector $X_{FO2,t} = (1, y_{t-1}, ISM_t^{(1)}, EMPLOY_t^{(1)})'$ has only four regressors.[5] Observe that for the CCM-LRG model the dimension of the $X_{m,t}$ vector is 34, 10, 22, and 34 for each forecast origin FO1–FO4. As discussed above, for a given sample size typical in the macroeconometric literature, OLS estimation of these multiple-indicator U-MIDAS models is likely to have adverse effects on model forecasting performance. Indeed, Carriero et al. (2015, p. 844) mention significant worsening in forecasting accuracy of the models estimated without shrinkage compared to those with Bayesian shrinkage.

4. EVALUATION METRICS

The evaluation metrics we use for the assessment of forecast accuracy of the individual models are the root mean squared forecast error (RMSFE) for point forecasts:

$$RMSFE = \sqrt{\frac{\sum_{t=\underline{\tau}}^{\overline{\tau}} (y_t - \widehat{y}_t)^2}{T}} \tag{4}$$

and the average logarithmic score (ALS) for density forecasts:

$$ALS = \frac{\sum_{t=\underline{\tau}}^{\overline{\tau}} LS(y_t)}{T} \tag{5}$$

where T is the number of observations in the forecast evaluation sample $[\underline{\tau}, \overline{\tau}]$, an actual outturn is denoted y_t and the model forecast is denoted $\widehat{y}_t$. The logarithmic score is defined as the negative logarithm of the value of the predictive density at the outturn

$LS(y_t) = -\ln F_t(y_t)$ (6),

following Gneiting and Katzfuss (2014). Since these evaluation metrics do not depend on the forecast origin, we omitted the corresponding subindex m.

For the pairwise comparison of models' forecast accuracy we use the Relative RMSFE (RRMSFE) and the average logarithmic score difference (ALSD) for point and

[4] Observe that at each forecast origin lagged values of the dependent variable, y_{t-i}, belong to different GDP data vintages. For the sake of notational simplicity we have suppressed this notation.
[5] The consequences of including higher lags of skip-sampled monthly variables are investigated in the working paper version of Carriero et al. (2015); see Carriero et al. (2013). It is reported there that their inclusion did not affect the results much compared to more parsimonious models.

density forecasts respectively. The RRMSFE and ALSD are defined as follows for each pair of models 1 and 2:

$$RRMSFE_{2/1} = \frac{RMSFE_2 - RMSFE_1}{RMSFE_1} \tag{7}$$

$$ALSD_{2/1} = ALS_2 - ALS_1 \tag{8}.$$

Since both the RMSFE and ALSD are negatively oriented scores, meaning that their larger values indicate less precise point and density forecasts, their positive/negative values indicate that on average the forecasting performance of model 2 is worse/better than that of model 1.

In addition to evaluation metrics based on the forecasting performance averaged over the forecast evaluation sample, we use the following measures of forecast accuracy: the cumulated sum of squared forecast error difference (CSSFED) of Welch and Goyal (2008) and the cumulated sum of logarithmic score difference (CSLSD), also known as the cumulative log predictive Bayesian factors for recursive evaluation of point and density forecasts, respectively. While the advantages of using recursive metrics for comparison of models' forecast densities were already highlighted in Geweke and Amisano (2010, p. 220) stating that this way of model comparison "(..) shows how individual observations contribute to the evidence in favor of one model over another. For example, it may show that a few observations are pivotal in the evidence strongly favoring one model over another", the use of such recursive metrics as the CSSFED for comparison of point forecasts is still rare in the macroeconomic forecasting literature.

Denoting by $e_{1,t}$ and $e_{2,t}$ the forecast errors of models 1 and 2 in period t, the CSSFED is computed as follows

$$CSSFED_{[\underline{\tau},\overline{\tau}],2/1} = \textstyle\sum_{s=\underline{\tau}}^{t} \left(e_{2,s}^2 - e_{1,s}^2\right) \quad for \quad t \in \left[\underline{\tau}, \overline{\tau}\right] \tag{9}$$

where $\left[\underline{\tau}, \overline{\tau}\right]$ denotes the forecast evaluation sample. The CSLSD is defined as follows:

$$CSLSD_{[\underline{\tau},\overline{\tau}],2/1} = \textstyle\sum_{s=\underline{\tau}}^{t} \left(LS_2(y_s) - LS_1(y_s)\right) \quad for \quad t \in \left[\underline{\tau}, \overline{\tau}\right] \tag{10}.$$

The recursive measures of forecast accuracy dissect the forecasting performance of the models observation by observation, illustrating how the relative forecasting performance evolves over time. As a result, both the CSSFED and CSLSD deliver a sequence of cumulative differentials, as opposite to aggregated measures of forecasting performance that deliver a point estimate. The recursive measures are helpful in distinguishing sources of domination of one model over its competitor in terms of forecasting accuracy. For example, a steady increasing $CSSFED_{2/1}$ indicates that the forecasting gains of model 1 over model 2 slowly accrue over time, i.e. the squared errors of model 2 tend to be marginally but systematically *larger* than those of model 1. Correspondingly, an upwards trending $CSLSD_{2/1}$ would indicate that the log-scores of model 1 tend to be marginally but systematically *smaller* than those of model 2, favouring the former. By the same token, a steady decreasing sequence of CSSFED/CSLSD indicates the opposite. A CSSFED/CSLSD sequence that evolves horizontally indicates that none of the competing models produces more accurate forecasts in a systematic manner.

CSSFED/CSLSD may also display abrupt jumps, indicating that in a given period t^* the differential of squared forecast errors/log-scores is substantially larger than those observed for the most of the observations. From the point of view of a forecasting practitioner, it is of special interest to detect such periods when one model suddenly displays worsening in its forecasting accuracy relative to its competitor. When comparing models in terms of their average forecasting performance, these periods turn out to be very influential for relative model ranking.

5. RESULTS

5.1 Point forecasts

Table 2 contains the results of point forecasting performance of the benchmark and the multiple-indicator models with constant volatility. We report the RMSFE and RRMSFE for the whole sample as well as separately for its expansionary and recessionary sub-periods. Observe that the forecasting accuracy reported for CCM-SML and CCM-LRG with constant volatility for the full sample in the upper panel of Table 2 is comparable with that reported for the corresponding models *Small BMF* and *Large BMF* in Table 2 of Carriero et al. (2015, p. 849). As in the original study, we notice that the RMSFE steadily decreases both in absolute value and value relative to the benchmark AR(2) model, as more information is incorporated into the indicator-augmented forecasting models. For example, as measured by the RRMSFE, the multiple-indicator models bring about 7% and 21%–22% gains in terms of RMSFE relative to the benchmark model at the first (FO1) and last (FO4) forecast origins, respectively.

We, however, are interested in determining whether such encouraging evidence in favour of the multiple-indicator models remains if we split the forecast evaluation sample into recessionary and expansionary sub-periods. The evidence reported in the middle and lower panels of Table 2 for the boom and bust periods suggests that this is not the case. First, we observe a large discrepancy between business cycle phases in the reported RMSFEs of the benchmark model. The RMSFEs in the recessionary periods are more than twice as large as in the expansionary periods.

Table 2
Point forecast accuracy

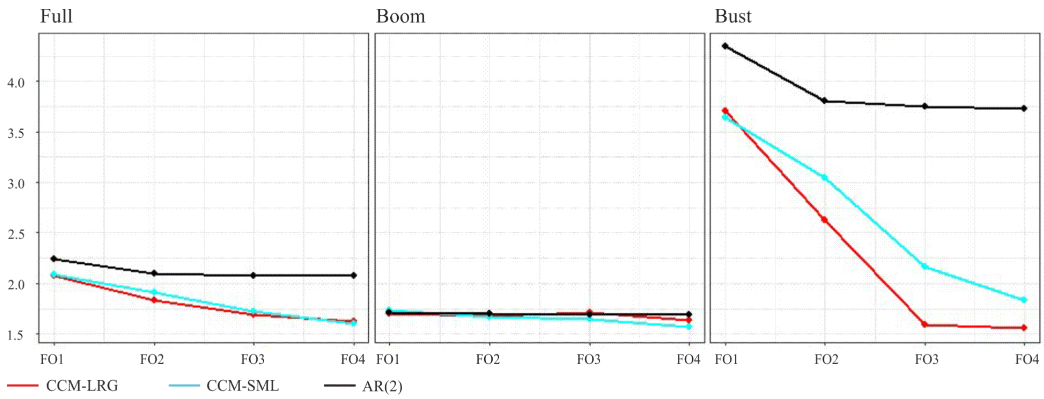
	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$RRMSFE_{\bullet/AR(2)}$*			
Full sample								
AR(2)	2.237	2.102	2.081	2.074				
CCM-SML	2.082	1.906	1.723	1.605	–0.069	–0.093	–0.172	–0.226
CCM-LRG	2.074	1.831	1.693	1.622	–0.073	–0.129	–0.186	–0.218
Boom sample								
AR(2)	1.710	1.705	1.692	1.691				
CCM-SML	1.731	1.669	1.647	1.567	0.013	–0.021	–0.027	–0.073
CCM-LRG	1.696	1.678	1.708	1.631	–0.008	–0.016	0.009	–0.035
Bust sample								
AR(2)	4.339	3.802	3.751	3.72				
CCM-SML	3.636	3.042	2.160	1.835	–0.162	–0.200	–0.424	–0.508
CCM-LRG	3.709	2.630	1.590	1.556	–0.145	–0.308	–0.576	–0.582

* The symbol • in $RRMSFE_{\bullet/AR(2)}$ denotes the name of a competing multiple-indicator model, i.e. CCM-SML or CCM-LRG.

Second, for the multiple-indicator models we observe a similar situation, but the difference between the phase-specific RMSFEs is narrowing with the forecast origin for CCM-SML and CCM-LRG for all FO1–FO4 and disappears for CCM-LRG at FO3–FO4. This finding conforms with that of Chauvet and Potter (2013) that the forecasting ability of macroeconometric models significantly worsens during recessions. However, we find that at least for the models in question this is true for models based on the partial information set, i.e. the univariate benchmark model, the CCM-SML model that uses only five out of 12 available indicators and the CCM-LRG model with all the 12 indicators but for the earliest forecast origins FO1–FO2.

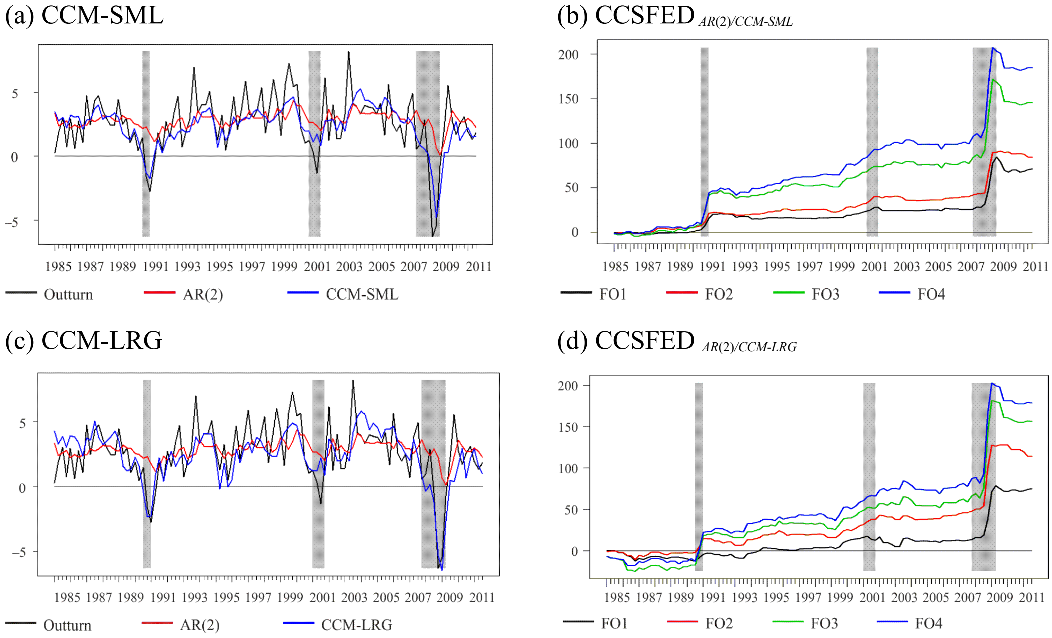
Third, the asymmetry in the forecasting ability of the indicator-augmented models is even more pronounced when measured in the relative terms with respect to the benchmark model. During expansions at the earlier forecast origins FO1–FO3, there is a barely noticeable difference as measured by the relative RMSFE that takes values in the range from –2.7% to 1.3%. It is only for FO4 that the RRMSFE reaches –7.3% for CCM-SML and –3.5% for CCM-LRG, implying substantially lower improvements over the benchmark model than recorded for the whole sample. At the same time, during the NBER recessions there is a very noticeable improvement in the forecasting ability of the indicator-augmented models compared to the benchmark model. The RRMSFE is about –15% at FO1, reaching impressive –50.8% and –58.2% for CCM-SML and CCM-LRG. The information on the RMSFEs reported in Table 2 is visually presented in Figure 4.

Figure 4
RMSFE summary



The actual outturns and forecast values for CCM-SML and CCM-LRG at FO4 are shown in Figures 5(a) and 5(c). The corresponding CSSFEDs of the indicator-augmented models with respect to the benchmark model are displayed in Figures 5(b) and 5(d) for all forecast origins. Examination of the plotted CSSFEDs exposes the causes of the asymmetry in forecasting performance as measured by the RRMSFEs in Table 2. The jumps in the CSSFEDs observed during recessions indicate that during the shaded periods the benchmark model produces much larger (squared) forecast errors than the multiple-indicator models. During the expansionary periods, the CSSFEDs display largely horizontal movements, indicating that none of the competing models systematically exhibits greater forecasting accuracy.

Figure 5
Recursive assessment of point forecast accuracy; CCM-SML/CCM-LRG vs AR(2)



5.2 Density forecasts

The results of the evaluation of the performance of the models in terms of density forecasts are summarised in Table 3. In line with Carriero et al. (2015), density forecasts were generated from the indicator-augmented and benchmark models with stochastic volatility. Observe that for the full sample these results are very close to those reported for the models *Small BMFSV* and *Large BMFSV* in Carriero et al. (2015, Table 3, p. 852).

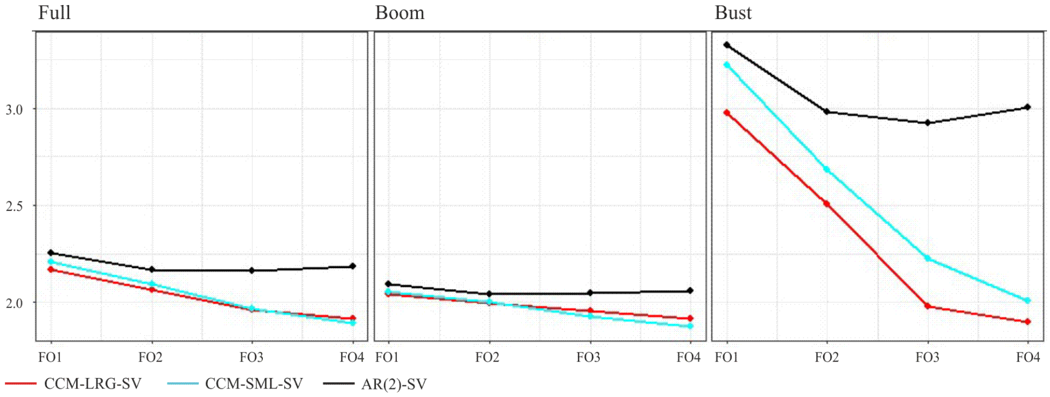
In addition, there are interesting parallels that can be drawn when comparing density forecasts with point forecasts reported in Sub-section 5.1. First, there is a large discrepancy in the quality of density forecasts generated by the benchmark model. In recessions the reported ALSs are much higher than in expansions. Second, we observe a similar pattern for the multiple-indicator models, but the differences in density forecast accuracy are more pronounced at the earlier forecast origins. For both indicator-augmented models the difference narrows steadily with each forecast origin, and for the CCM-LRG-SV model it practically disappears as early as at FO3. Third, the improvement in relative forecast accuracy of multiple-indicator models over the benchmark model measured in terms of ALSD steadily increases. For example, for the CCM-LRG-SV the ALSD is reported as –0.347 at FO1, which is to be compared with the corresponding value of –1.107 at FO4. The information on ALSs reported in Table 3 is visually presented in Figure 6.

Table 3
Density forecast accuracy

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				ALSD$_{\bullet/AR(2)\text{–}SV}$*			
Full sample								
AR(2)-SV	2.253	2.166	2.159	2.182				
CCM-SML-SV	2.207	2.090	1.965	1.891	–0.046	–0.076	–0.195	–0.292
CCM-LRG-SV	2.165	2.063	1.959	1.913	–0.088	–0.103	–0.200	–0.270
Boom sample								
AR(2)-SV	2.092	2.043	2.044	2.059				
CCM-SML-SV	2.054	2.001	1.926	1.873	–0.038	–0.042	–0.118	–0.185
CCM-LRG-SV	2.043	1.996	1.956	1.915	–0.049	–0.047	–0.088	–0.144
Bust sample								
AR(2)-SV	3.324	2.982	2.926	3.004				
CCM-SML-SV	3.223	2.682	2.223	2.006	–0.101	–0.300	–0.703	–0.998
CCM-LRG-SV	2.977	2.508	1.976	1.898	–0.347	–0.474	–0.950	–1.107

* The symbol • in ALSD$_{\bullet/AR(2)\text{-}SV}$ denotes the name of a competing multiple-indicator model, i.e. CCM-SML-SV or CCM-LRG-SV.

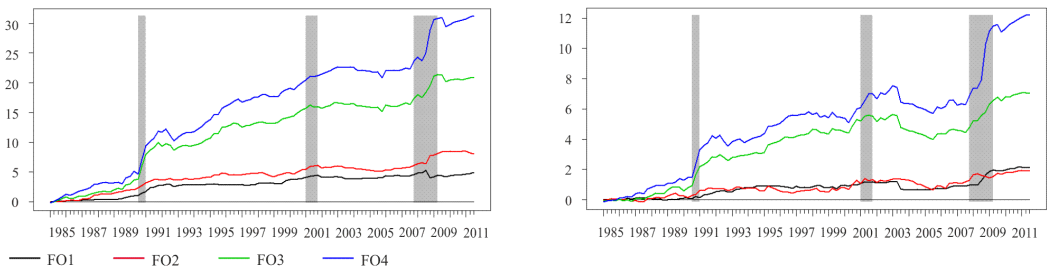
Figure 6
ALS summary



Additional information on the relative forecasting performance of the models can be obtained from the CSLSDs displayed in Figure 7. The observed jumps during the shaded periods reveal the main sources of the relatively better forecasting performance of indicator-augmented models in comparison to univariate benchmark model.

Figure 7
Recursive assessment of density forecast accuracy; CCM-SML-SV/CCM-LRG-SV vs AR(2)-SV

(a) CSLSD $_{AR(2)\text{-}SV/CCM\text{-}SML\text{-}SV}$ (b) CSLSD $_{AR(2)\text{-}SV/CCM\text{-}LRG\text{-}SV}$



All in all, the conclusions on the importance of distinguishing between business cycle phases when evaluating models in terms of their out-of-sample accuracy made for point forecasts are equally valid for density forecasts. The consequences of ignoring it are also qualitatively similar: reporting ALSD for the whole period generally biases the evaluation of the relative forecasting performance of the models in favour of indicator-augmented models during expansions and consequently understates their predictive ability during recessions.

5.3 Impact of stochastic volatility

In this sub-section we reassess the consequences on accuracy of point and density forecasts from introducing stochastic volatility into multiple-indicator models. For the full sample, this was already done in Carriero et al. (2015). They conclude that though there is no noticeable effect on the accuracy of point forecasts, adding stochastic volatility to the models greatly enhances the accuracy of density forecasts – a conclusion already established in related research (Clark (2011)).

It is of interest to verify whether these conclusions hold if one compares the point and density forecasting performance of multiple-indicator models with and without stochastic volatility separately for expansions and recessions, similarly as we have done above when we compared their performance with the benchmark model.

The effect of stochastic volatility on point and density forecast accuracy is reported in Table 4. The upper panel reports the values of $\text{RRMSFE}_{\bullet - SV/\bullet}$ with • = *CCM-SML, CCM-LRG* for the full sample as well as for expansionary and recessionary sub-periods. An overall impression is that adding stochastic volatility to the model, when the purpose is generation of point forecasts, does not result in systematic improvements. This finding appears to hold both when evaluated for the whole sample and across business cycle phases. Depending on the forecast origin and the model (CCM-LRG or CCM-SML), the RRMSFE varies in the range from –0.053 to 0.107.

Table 4
Impact of stochastic volatility on forecast accuracy

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	Full sample				Boom sample				Bust sample			
Point forecasts: $\text{RRMSFE}_{\bullet - SV/\bullet}$												
CCM-SML	0.011	0.007	0.000	–0.019	0.001	0.002	–0.011	–0.024	0.027	0.018	0.043	0.001
CCM-LRG	–0.016	0.009	0.005	–0.025	0.010	0.010	–0.009	–0.029	–0.053	0.007	0.107	0.007
Density forecasts: $\text{ALSD}_{\bullet - SV/\bullet}$												
CCM-SML	–0.194	–0.215	–0.197	–0.214	–0.296	–0.265	–0.222	–0.224	0.479	0.121	–0.034	–0.147
CCM-LRG	–0.203	–0.184	–0.133	–0.134	–0.264	–0.227	–0.140	–0.133	0.201	0.105	–0.087	–0.141

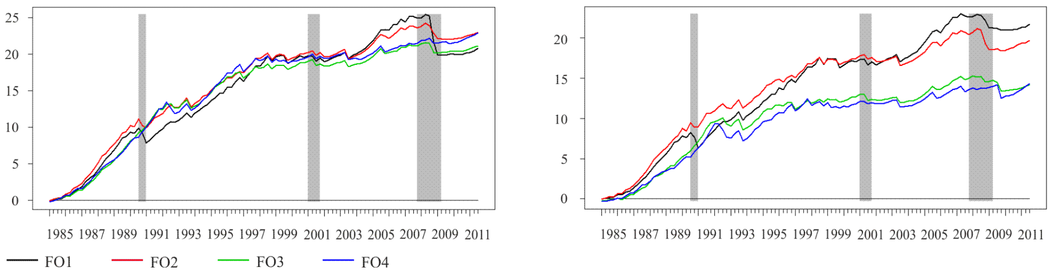
The symbol • in $\text{RRMSFE}_{\bullet - SV/\bullet}$ and $\text{ALSD}_{\bullet - SV/\bullet}$ denote the name of a multiple-indicator model, i.e. either CCM-SML or CCM-LRG.

A more interesting pattern emerges when one analyses the effect of stochastic volatility on the accuracy of density forecasts (see the lower panel of Table 4). The entries in the panel are differences in ALS between models with stochastic and constant volatility in the residual error term, $\text{ALSD}_{\bullet - SV/\bullet}$. Negative entries imply that on average density forecast accuracy of the models with stochastic volatility is higher, positive entries indicate the opposite. As can be seen, for the full sample as well as for expansions the models with stochastic volatility produce lower ALSs than their counterparts with constant volatility. This is in line with the results reported in Carriero et al. (2015). However, for recessions the situation is different. For both models (CCM-SML and CCM-LRG) there is a positive difference in ALS at FO1–FO2, implying that the model versions with stochastic volatility produce inferior density forecasts than their homoscedastic counterparts. At the later forecast origins, FO3–FO4, this difference is negative.

In tracking the reason for this at the first glance surprising result, it is instructive to inspect plots of CSLSD for models with stochastic and constant volatility. These are displayed in Figure 8. In general, an upwards trending behaviour of the CSLSDs indicates more or less steady gains in density forecast accuracy of models with stochastic volatility over those with constant volatility. This behaviour is to be expected as the models with contant volatility cannot capture the reduction in the unconditional volatility of GDP growth since 1985, i.e. in the Great Moderation period. As a result, models with constant volatility tend to produce predictive densities that are too wide in comparison with those produced by models with stochastic volatility that are able to accommodate a decrease in the GDP volatility, and hence reduce the forecast uncertainty of GDP growth in the current quarter.

Figure 8
Impact of stochastic volatility on density forecast accuracy

(a) CSLSD $_{CCM\text{-}SML/CCM\text{-}SML\text{-}SV}$ (b) CSLSD $_{CCM\text{-}LRG/CCM\text{-}LRG\text{-}SV}$

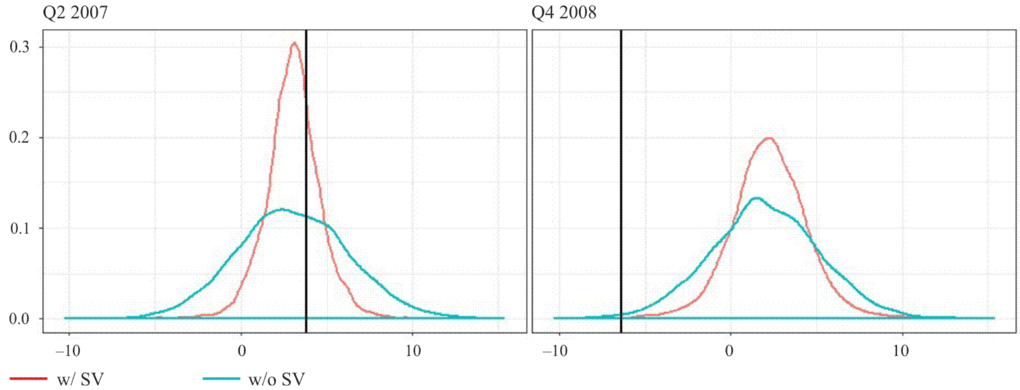


However, there are several observations when the latter models produced more accurate density forecasts. These observations belong to the recession period in the early 1990s, and especially the Great Recession.

For the CSLSD$_{CCM\text{-}SML/CCM\text{-}SML\text{-}SV}$ the worse performance is visible at all forecast origins during the Great Recession, whereas for the CSLSD$_{CCM\text{-}LRG/CCM\text{-}LRG\text{-}SV}$ it occurs at FO1–FO2 during the Great Recession and FO3–FO4 in the aftermath of the Great Recession. This points towards inferior predictive densities of SV-augmented models during the post-recession recovery period.

An explanation for this observation based on inspection of the CSLSD can be given using Figure 9, where the density forecasts from CCM-SML and CCM-SML-SV at FO1 are displayed for one representative quarter for the expansionary periods (the second quarter of 2007) and recessionary periods (the fourth quarter of 2008). As discussed above, the model with constant volatility (the blue line) produces too wide predictive densities in comparison with the model with stochastic volatility (the red line). The black vertical line denotes the outturn of the GDP growth rate in the respective quarter. During the representative expansionary quarter (the second quarter of 2007), the model with SV produces lower log score than the model without SV.[6] As shown in Figure 8, producing upward-sloping CSLSDs appears to be the rule rather than an exception. However, during the recessionary quarter (the fourth quarter of 2008), when the actual value lies much farther in the negative area than it used to be in normal times, the curse of the longer tail of the model with constant volatility turns into a blessing, resulting in a lower log score value than that produced by the model with time-varying variance of the error term.

Figure 9
Predictive densities for CCM-SML model at FO1



The following lesson can be learned from this exercise of comparing density forecasts from models with constant and time-varying volatility of the error term. For the data at hand introducing stochastic volatility brings about more accurate density forecasts most of the times. However, there might be several outturns that lie farther out in the tails, especially, during periods of economic downturns and, possibly, during the post-recession recoveries, when fatter tails of models with constant volatility may help in producing relatively lower log scores. In this respect it is worthwhile pointing out that models with stochastic volatility that were superior in capturing nowcast uncertainty most of the times failed to do so during crisis periods when there is stronger than ever demand for accurate assessment of economic conditions and risks involved.

As a final word, we used the additional example presented in this section in order to illustrate the benefits of recursive measures of models' relative forecast accuracy. We were able to uncover peculiarities in the relative forecasting performance of the models that remain concealed to Carriero et al. (2015) who relied on traditional evaluation methods.

6. CONCLUSIONS

In this paper we have re-assessed the forecasting performance of the model suggested in Carriero et al. (2015). The model is characterised by a number of practical features like dealing with mixed-frequency data, stochastic volatility and the Bayesian approach to estimation and generation of both point and density forecasts that make it highly appealing as a workhorse model at policy-making institutions such as central banks which continuously engage in nowcasting and forecasting activity.

Our research is motivated by the recent study of Chauvet and Potter (2013) that documented asymmetric performance of several popular state-of-the-art structural and

[6] Recall from equation (6) that the log score is computed as the negative (logarithmic) value of the predictive density at the realised outturn value, i.e. the intersection of the vertical black line with the blue and red lines for models with constant and stochastic volatility, respectively.

reduced-form macroeconometric models during expansions and recessions. Chauvet and Potter (2013) report their findings for models estimated using macroeconomic data sampled at a single (quarterly) frequency. In this paper we extend their analysis to forecasting models that utilise mixed-frequency data sampled both at the quarterly frequency such as GDP growth and monthly frequency such as economic and financial indicators like surveys, labour-market statistics, industrial production and interest rates. Such models utilise the informational flow within each quarter and hence allow making forecasts more than once per quarter in comparison to models estimated with only quarterly data.

Our main finding is that the conclusions of Chauvet and Potter (2013) hold also in our study. We record that during expansions the forecasting ability of more sophisticated models that rely on a larger information set are comparable to the forecast accuracy of benchmark univariate AR(2) model. It is only during recessions that indicator-augmented models substantially improve upon the benchmark model.

Our results have important implications when reporting the results of forecasting competitions by averaging them over the whole forecast sample typically involving both recessions and expansions. Failure to acknowledge business cycle asymmetries in the forecast accuracy of a more sophisticated forecasting model relative to benchmark models typically results in exaggeration of the relative forecasting accuracy of the former in expansions and, consequently, understatement during recessions. This delivers a biased message to anyone interested in forecasts – be it the general public, academics, practitioners or policy-makers.

In addition to evaluation of the relative forecasting performance separately for expansions and recessions, we use recursively computed metrics in order to illustrate our findings on the differences in relative forecasting ability of models across the business cycle phases. We used these recursively computed measures in order to uncover previously unnoticed features of forecasts from models with time-varying variance of the error term. Previous research widely documents that when dealing with US data, inclusion of stochastic volatility typically brings about more accurate density forecasts. However, we discover that models with stochastic volatility may sometimes underestimate the uncertainty around forecasts in comparison with models with constant volatility. In our study, we document this outcome in situations when it is least desired and expected, i.e. during crisis or post-crisis recovery periods when outcomes lie farther out in the tails of the GDP growth distribution.

As a final word, in our study we examined the forecasting performance of the model suggested in Carriero et al. (2015). By focusing only on one model type we were able to deliver a clear message on how our results compare with those of Carriero et al. (2015) serving as a well-documented benchmark. Undoubtedly, our analysis can be extended to other types of models such as mixed-frequency VARs (Schorfheide and Song (2015), McCracken et al. (2015), Mikosch and Neuwirth (2015)) though none of those incorporates stochastic volatility, mixed-frequency factor models (Marcellino et al. (2016), Marcellino and Schumacher (2010)) as well as models featuring regime-switching behaviour like a mixed-frequency version of the AR-dynamic factor model with Markov switching that fared very well in Chauvet and Potter (2013). An additional avenue for future research is to consider more sophisticated models for stochastic volatility proposed in Chan (2017) and Zhang et al. (2018). Extending our analysis to other data than US GDP seems also a fruitful research direction.

APPENDIX

In the Appendix we provide additional results on the asymmetric forecasting performance of econometric models across the business cycle phases. In Section A we address the question of chosing the benchmark model. To this end, we compare two popular choices for benchmark models in the forecasting literature: the AR(2) and historical-mean or RW models.

Next, we are interested in comparing two approaches to data aggregation: pooling many indicators into one model and pooling many SIMs by resorting to certain model combination schemes commonly used in the literature. The former approach was undertaken in the main text and the latter approach was used in Mazzi et al. (2014) when focusing specifically on density forecast combinations in mixed-frequency models. We intend to provide empirical evidence on this question using the data set at hand and explicitly comparing the model-pooling approach of Mazzi et al. (2014) with the data-pooling approach of Carriero et al. (2015). To this end, we evaluate forecasting performance of models augmented only with one indicator at a time – the so-called SIMs – and then we evaluate forecasting performance of combinations of these SIMs both in terms of point and density forecasts. The results of this exercise are reported in Section B for the SIMs and in Section C for their combinations. In Section D we evaluate the impact of stochastic volatility on the accuracy of point and density forecasts for these additional models.

Additional models

For the two benchmark models the conditional mean in equation (1) is specified as follows: for the AR(2) model the vector $X_{m,t}$ is $X_{m,t} = (1, y_{t-2}, y_{t-3})'$ for $m =$ FO1 and $X_{m,t} = (1, y_{t-1}, y_{t-2})'$ for $m =$ FO2,...,FO4; for the RW model the vector $X_{m,t} = (1)$ consists only of the intercept at all forecast origins.

The SIMs are constructed by allowing one monthly indicator at a time to be selected as a regressor. In sequel, we refer to such models as SIM-• where the symbol • stands for an abbreviation used for each of the monthly indicators (see Table 1). For such models the conditional mean at each forecast horizon is either:

FO1: $X_{m,t} = (1, y_{t-2}, w^{(1)}_{1,t-1}, w^{(2)}_{1,t-1}, w^{(3)}_{1,t-1})'$,

FO2: $X_{m,t} = (1, y_{t-1}, w^{(1)}_{1,t})'$,

FO3: $X_{m,t} = (1, y_{t-1}, w^{(1)}_{1,t}, w^{(2)}_{1,t})'$,

FO4: $X_{m,t} = (1, y_{t-1}, w^{(1)}_{1,t}, w^{(2)}_{1,t}, w^{(3)}_{1,t})'$

if an indicator is released during the first week of each month or

FO1: $X_{m,t} = (1, y_{t-2}, w^{(1)}_{2,t-1}, w^{(2)}_{2,t-1})'$,

FO2: $X_{m,t} = ()$,

FO3: $X_{m,t} = (1, y_{t-1}, w^{(1)}_{2,t})'$,

FO4: $X_{m,t} = (1, y_{t-1}, w^{(1)}_{2,t}, w^{(2)}_{2,t})'$

if a variable is released during the second week of each month. Observe that for this group of variables none of the contemporaneous values for the current quarter t is released at the second forecast origin, FO2. Therefore, similarly to Carriero et al. (2015), we do not estimate any model with any of these variables at FO2.

Furthermore, we use the SIMs in order to form their model combinations based on their past point and density forecasting performance. For model combinations based on the past point forecast performance we use (a) an equal weighting scheme as well as (b) time-varying weights that are recursively determined based on a discounted mean squared forecast error (MSFE) $\lambda_{m,t,i} = \sum_{s=\underline{\tau}}^{t-h_m} \delta^{t-h_m-s}(\hat{e}_{m,s,i})^2$, with the discounting factor $0 < \delta \leq 1$. h_m is the delay in quarters that is the specific of every forecast origin indicating availability of the second release of GDP data allowing us to calculate forecast errors at the forecast origin in question: for FO1/FO2 and FO3/FO4 the delay parameter is $h_m = 2$ and $h_m = 1$. The combination weights are computed using the following expression:

$$\mathcal{W}_{m,t,i} = \frac{\lambda_{m,t,i}^{-1}}{\sum_{j=1}^{n} \lambda_{m,t,j}^{-1}} \quad (11).$$

We initialise the weighting scheme by using equal weights for all estimated models at a given forecast origin.

For model combinations based on their past density forecasting performance we follow Mazzi et al. (2014) in using the linear opinion pool approach:

$$LP_{m,t} = \sum_{j=1}^{n} \mathcal{W}_{m,t,j}\mathcal{F}_{m,t,j} \quad for \quad t \in (\underline{\tau}, \overline{\tau})$$

where $\mathcal{F}_{m,t,k}$ is the nowcast density from a model $k = 1,\ldots,n$ for quarter t. As above for point forecasts, we set the weights in two ways (a) by using equal weights for all model nowcast densities and (b) by using recursive weights. Denoting by $\ln\mathcal{F}_{m,t,j}(y_t)$ the logarithm of the value of the nowcast density at the outturn y_t, the recursive weights are determined as follows:

$$\mathcal{W}_{m,t,k} = \frac{\exp\left[\sum_{s=\underline{\tau}}^{t-h_m} \ln\mathcal{F}_{m,s,k}(y_s)\right]}{\sum_{j=1}^{n} \exp\left[\sum_{s=\underline{\tau}}^{t-h_m} \ln\mathcal{F}_{m,s,j}(y_s)\right]} \quad (12)$$

where h_m takes values of one and two for FO1–FO2 and FO3–FO4, respectively, as in the case of point forecast combination. We initialise the weighting scheme by using equal weights for all estimated models at a given forecast horizon.

We denote model combinations based on point forecasting performance with equal weighting by CPF-EW and with recursively defined weights by CPF-RW100, CPF-RW090 and CPF-RW030, each corresponding to the following values of the discounting factor δ = 1,0.90,0.30. The model combinations based on density forecasting performance are denoted by CDF-EW and CDF-RW for those determined by equally and recursively defined weights.

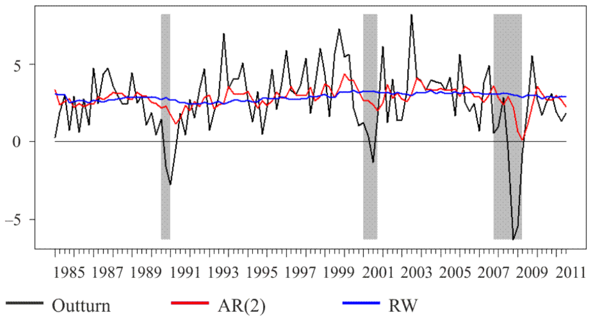
A. Comparative performance of the benchmark models

A.1 Point forecasts

When comparing models in terms of point forecast accuracy, Carriero et al. (2015) used a univariate AR(2) model with constant volatility as a benchmark model. In this sub-section we conduct an additional analysis for the choice of the benchmark model by comparing the forecasting performance of the AR(2) model with that of the RW model.

The actual values of GDP growth together with forecasts produced by the AR(2) and RW models are shown in Figure A.1.[7] The forecasts of the AR(2) model fluctuate around those of the RW model. Notably, it is easy to recognise that in times of recessions the AR(2) model produces more accurate forecasts as its forecasts tend to be below those of the RW model and hence much closer to the outturns of GDP growth. At the same time, during expansions it is not that obvious that the former model on average tends to produce more accurate forecasts than the latter one. In order to sort this out, we resort to the formal statistical analysis below.

Figure A.1
Outturns and point forecasts from the benchmark AR(2) and RW models



Further information on the comparative forecasting performance of these two models is provided in the left and right panels of Figure A.2. In the left panel squared forecast errors for both models are shown making it evident that in times of recessions their magnitude tends to be larger than in times of expansions. In the right panel we report differences in the squared forecast errors which also tend to be larger in recessions than in expansions. This finding may imply that differences in the overall forecasting performance measured by the RMSFE for the whole forecasting sample are mainly driven by the observations pertaining to recessions rather than to expansions. We will verify this idea formally below by comparing forecasting accuracy of these models for the full forecast evaluation sample as well as for its boom and bust sub-samples.

Additional information on relative forecasting performance in terms of point forecasts is provided by the CSSFED displayed in Figure A.3. In expansionary periods the CSSFED mostly displays flat horizontal movements. As discussed in Section 4, the dynamics conform with the idea that during expansions there is no systematic difference between forecast accuracy of these two models. In contrast, during recessions there are noticeable jumps in the CSSFED indicating that the RW model produces much larger squared forecast errors than the AR(2) one.

[7] In order to keep the discussion concise, we limit it to forecasting results reported for the forecast origin FO4. The results obtained for the rest of the forecast origins are qualitatively very similar.

Figure A.2

Squared point forecast errors from the AR(2) and RW models and their difference

(a) Squared forecast errors, AR(2) and RW models

(b) Squared forecast error difference, $FE^2_{RW,t} - FE^2_{AR(2),t}$

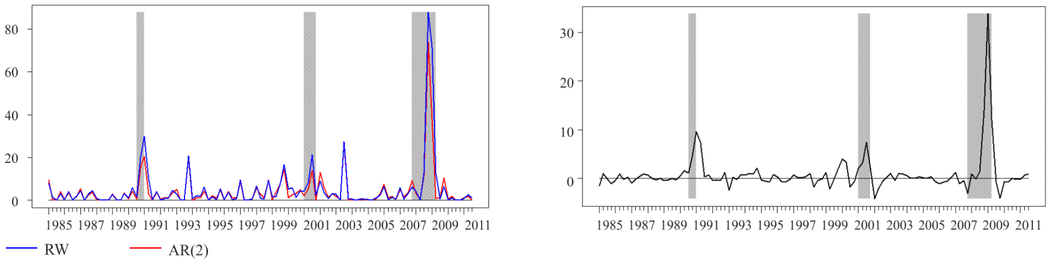
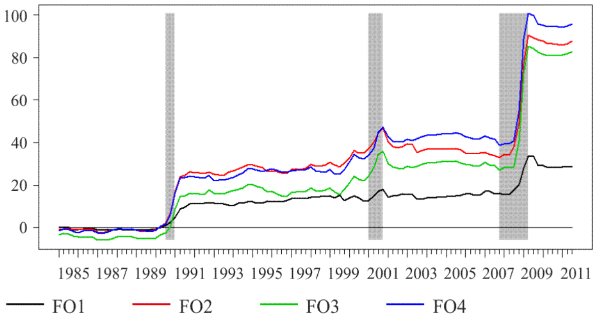


Figure A.3

Point forecast accuracy, CSSFED$_{RW/AR(2)}$



The results of the formal comparison in terms of RMSFEs and RRMSFEs are reported in Table A.1 for the full forecast sample as well as for its boom and bust sub-samples. Several interesting observations can be made. First, both models display much larger average values of (squared) forecast errors during recessions than expansions. This means that reporting RMSFE for the full sample understates forecast accuracy of the models in question during expansions and overstates it during recessions.

Table A.1

Point forecast accuracy, univariate benchmark models

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				RRMSFE$_{RW/AR(2)}$			
Full sample								
AR(2)	2.237	2.102	2.081	2.074				
RW	2.296	2.289	2.259	2.280	0.027	0.089	0.086	0.099
Boom sample								
AR(2)	1.710	1.705	1.692	1.691				
RW	1.713	1.716	1.695	1.712	0.002	0.006	0.002	0.013
Bust sample								
AR(2)	4.339	3.802	3.751	3.726				
RW	4.562	4.526	4.461	4.499	0.051	0.190	0.189	0.207

Second, during expansions the values of RMSFEs of the RW model are comparable to those of the AR(2) model. It is only during recessions when the difference between models' RMSFE values becomes non-negligible with the AR(2) model, delivering greater forecast accuracy. This result has implications for comparing models in terms of their forecasting accuracy for the whole sample. In doing so, one tends to exaggerate relative forecasting performance of the AR(2) model with respect to that of the RW model during expansions and, conversely, to understate it during recessions. For example, from the relative RMSFEs reported in the right panel of Table A.1 one can deduce that the RW model produces an RMSFE that is higher than that of the AR(2) model by up to 10% when these RMSFEs are computed over the whole forecast evaluation sample. At the same time, during booms the RW model produces RMSFE that is higher than that of the AR(2) model at most by about 1%. During busts the RW model produces an RMSFE that is higher than that of the AR(2) model up to about 20%.

Lastly, the fact that the AR(2) model excels over the RW model only during recessions and that both benchmark models produce very similar forecast accuracy during expansions implies that when comparing forecasting performance of the quarterly US GDP growth rate against the benchmark AR(2) model during expansions the comparison is effectively done against forecast accuracy of a historical mean model (Carriero et al. (2015), Chauvet and Potter (2013), *inter alia*).

A.2 Density forecasts

When comparing models in terms of density forecast accuracy, Carriero et al. (2015) used a univariate AR(2) model with stochastic volatility as a benchmark model. In this sub-section we conduct an additional analysis for the choice of the benchmark model by comparing the AR(2) and RW models in terms of density forecast accuracy. Average values of log scores and their difference for the models in question are presented in Table A.2. The asymmetry in the forecasting performance of the models across business cycle phases also manifests itself for density forecasts. Similarly, as in the case of point forecasts density forecasts are more accurate during expansions than during recessions when comparing the average log scores reported for each of the business cycle phases. This asymmetry is much more pronounced for the RW model. When comparing the relative forecasting performance of these two models, it is noticeable that the $ALSD_{RW\text{-}SV/AR(2)\text{-}SV}$ values are much more negative in recessions than in expansions. This indicates that the lion's share of the evidence in favour of the AR(2) model over the RW model stems from recession periods.

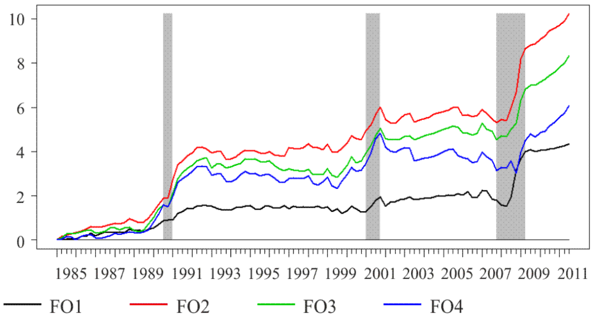
Table A.2
Density forecast accuracy, univariate benchmark models

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$ALSD_{RW\text{-}SV/AR(2)\text{-}SV}$			
Full sample								
AR(2)-SV	2.253	2.166	2.159	2.182				
RW-SV	2.294	2.261	2.237	2.239	0.041	0.096	0.078	0.057
Boom sample								
AR(2)-SV	2.092	2.043	2.044	2.059				
RW-SV	2.106	2.091	2.090	2.088	0.014	0.048	0.046	0.029
Bust sample								
AR(2)-SV	3.324	2.982	2.926	3.004				
RW-SV	3.543	3.393	3.218	3.245	0.219	0.412	0.292	0.241

Additional information on relative forecasting performance of univariate models in terms of density forecasts is provided by the CSLSD displayed in Figure A.4. The strongest increases in the CSLSD can be observed during recessions. In contrast, during the periods between recessions the CSLSD displays more heterogeneous dynamics: there are some periods with practically horizontal movements, periods characterised by upwards trends, especially shortly before or shortly after recessions, and some periods with (slightly) downwards trending dynamics. Such difference in the CSLSD dynamics during booms and busts explains differences in the ALS values reported in Table A.2 supporting the conclusion that most of the evidence driving differences in the forecasting accuracy of the competing models is due to observations during recessions.

Figure A.4

Denstity forecast accuracy, CSLS$_{RW\text{-}SV/AR(2)\text{-}SV}$



B. Single-Indicator Model

B.1 Point forecasts

The forecasting performance of SIMs against the benchmark AR(2) model is summarised in Table B.1.[8] Here again we report the RMSFEs and RRMSFEs against the benchmark model for the full sample as well as for its two sub-samples. A look at the reported RMSFEs in the table reveals that also for the SIMs there is a strong evidence of asymmetric forecasting performance during booms and busts. In absolute value the RMSFEs are higher during busts than booms for all SIMs. This finding is consistent with the statement of Chauvet and Potter (2013) that during recessions it is more difficult to forecast than during expansions. We also notice that the forecasting performance varies from one indicator to another. One can single out several indicators such as ISM, EMPLOY, ORDERS that provide much more accurate forecasts than others. The rest of indicators tend to be less informative both when their forecasting performance is evaluated either for the whole sample or for its sub-samples.

The heterogeneous results in reported RMSFE have also implications when one evaluates the forecasting performance of the SIMs relative to that of the benchmark AR(2) model (see the right panel of Table B.1). When evaluated for the whole sample, the three SIMs (ISM, EMPLOY, ORDERS) display the largest gains in the forecasting accuracy relative to the AR(2) among all 12 indicators. However, the comparison of the RRMSFE values in that table reveals that the most of improvement in the relative forecasting accuracy accrues during the recession periods. During expansions the benefits of using these SIMs over the benchmark model are less pronounced. For example, at FO4 the gains in RMSFE achieved by any of this three SIMs relative to the AR(2) model is about 5%–6% during expansions versus approximately 29%–36%

[8] The choice of the AR(2) model as a benchmark model is intentional to make our results directly comparable to those reported in Carriero et al. (2015).

during recessions. It is worthwhile mentioning that during recessions the RMSFE of the housing starts variable (STARTS) is comparable to that of ISM, EMPLOY, ORDERS variables. However, this is not the case during expansions when the forecasting accuracy of the SIM-STARTS model is the worst among the rest of the SIMs and even worse than that of the benchmark model, especially at FO4. The information presented in Table B.1 can be visually assessed using Figure B.1.

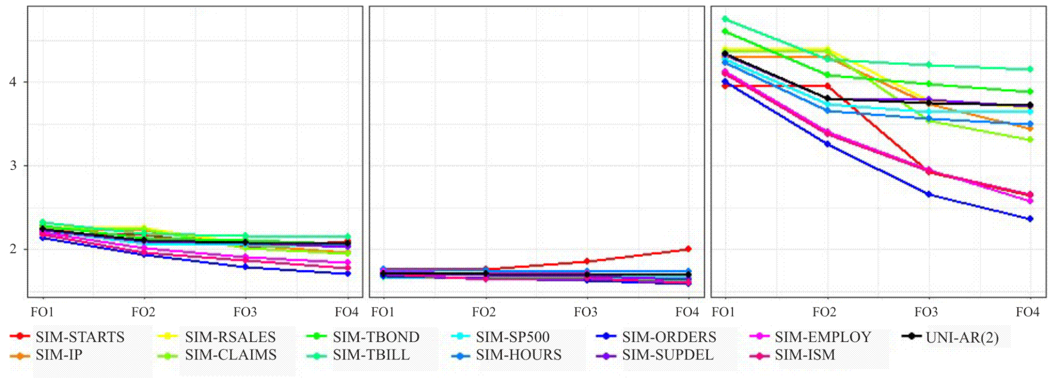
All in all, when reporting the results of forecasting accuracy of the SIMs without accounting for their differential performance during business cycle phases, the readers are misled into believing that the relative performance of indicator-augmented models in comparison with the benchmark model is better/worse during expansions/ recessions than it really is.

Table B.1
Point forecast accuracy, SIMs

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$RRMSFE_{RW/AR(2)}$*			
Full sample								
AR(2)	2.237	2.102	2.081	2.074				
SIM-ISM	2.172	1.960	1.870	1.775	–0.029	–0.068	–0.101	–0.144
SIM-EMPLOY	2.198	2.012	1.913	1.835	–0.017	–0.043	–0.081	–0.115
SIM-SUPDEL	2.234	2.094	2.081	2.031	–0.001	–0.004	0.000	–0.021
SIM-ORDERS	2.137	1.938	1.793	1.714	–0.045	–0.078	–0.138	–0.174
SIM-HOURS	2.242	2.099	2.068	2.057	0.002	–0.002	–0.006	–0.008
SIM-SP500	2.225	2.073	2.053	2.037	–0.005	–0.014	–0.013	–0.018
SIM-TBILL	2.316	2.186	2.167	2.144	0.035	0.040	0.041	0.034
SIM-TBOND	2.281	2.137	2.105	2.070	0.020	0.016	0.012	–0.002
SIM-CLAIMS	2.242		2.020	1.948	0.002		–0.029	–0.061
SIM-RSALES	2.249		2.075	2.042	0.006		–0.003	–0.015
SIM-IP	2.225		2.055	1.960	–0.005		–0.012	–0.055
SIM-STARTS	2.178		2.022	2.091	–0.026		–0.028	0.008
Boom sample								
AR(2)	1.710	1.705	1.692	1.691				
SIM–ISM	1.702	1.644	1.650	1.603	–0.005	–0.036	–0.025	–0.052
SIM-EMPLOY	1.730	1.708	1.702	1.696	0.012	0.001	0.006	0.003
SIM-SUPDEL	1.713	1.693	1.681	1.636	0.002	–0.008	–0.006	–0.032
SIM-ORDERS	1.684	1.651	1.622	1.592	–0.015	–0.032	–0.041	–0.058
SIM-HOURS	1.756	1.746	1.736	1.738	0.027	0.024	0.026	0.028
SIM-SP500	1.720	1.688	1.690	1.666	0.006	–0.010	–0.001	–0.015
SIM-TBILL	1.664	1.659	1.657	1.644	–0.027	–0.027	–0.020	–0.028
SIM-TBOND	1.670	1.656	1.649	1.633	–0.023	–0.029	–0.025	–0.034
SIM-CLAIMS	1.708		1.675	1.648	–0.001		–0.010	–0.025
SIM-RSALES	1.711		1.675	1.654	0.001		–0.010	–0.022
SIM-IP	1.709		1.662	1.624	–0.001		–0.018	–0.039
SIM-STARTS	1.763		1.848	1.995	0.031		0.092	0.180
Bust sample								
AR(2)	4.339	3.802	3.751	3.726				
SIM-ISM	4.101	3.377	2.939	2.650	–0.055	–0.112	–0.216	–0.289
SIM-EMPLOY	4.129	3.401	2.952	2.573	–0.048	–0.106	–0.213	–0.309
SIM-SUPDEL	4.319	3.805	3.787	3.705	–0.005	0.001	0.010	–0.006
SIM-ORDERS	4.009	3.255	2.662	2.368	–0.076	–0.144	–0.290	–0.365
SIM-HOURS	4.236	3.663	3.560	3.503	–0.024	–0.037	–0.051	–0.060
SIM-SP500	4.265	3.731	3.637	3.646	–0.017	–0.019	–0.030	–0.022
SIM-TBILL	4.755	4.272	4.198	4.145	0.096	0.124	0.119	0.112
SIM-TBOND	4.608	4.083	3.975	3.878	0.062	0.074	0.060	0.041
SIM-CLAIMS	4.365		3.541	3.309	0.006		–0.056	–0.112
SIM-RSALES	4.385		3.778	3.701	0.011		0.007	–0.007
SIM-IP	4.294		3.731	3.441	–0.010		–0.005	–0.077
SIM-STARTS	3.949		2.925	2.638	–0.090		–0.220	–0.292

* The symbol • in $RRMSFE_{\bullet/AR(2)}$ denotes the name of a competing SIM.

Figure B.1
RMSFE summary, SIMs



In order to illustrate the forecasting performance of the SIMs, we present actual and forecast values as well as the corresponding CSSFEDs for two selected models (SIM-ORDERS and SIM-STARTS) against the AR(2) model in Figure B.2 for FO4 and Figure B.3 for FO1, FO3–FO4, respectively. These two figures show that indeed most of the forecasting gains relative to the benchmark model accumulate during recessions.

Figure B.2
Actual outturns and point forecasts for selected SIMs

(a) SIM-ORDERS vs AR(2) (b) SIM-STARTS vs AR(2)

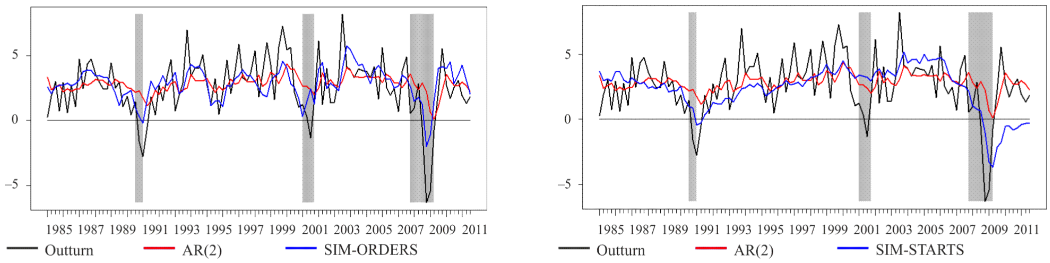
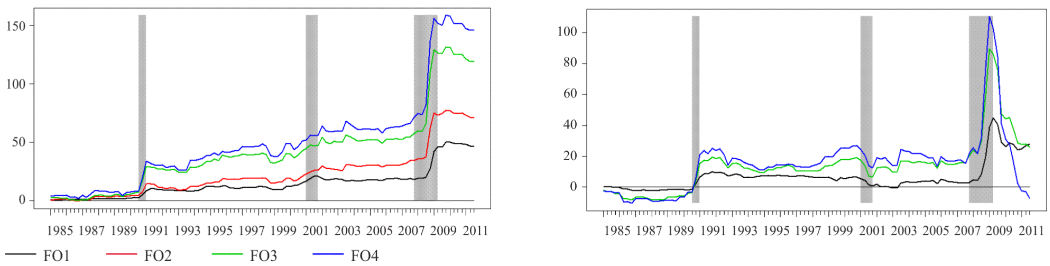


Figure B.3
CSSFED computed for selected SIMs

(a) CCSFED $_{\text{AR2/SIM-ORDERS}}$ (b) CCSFED $_{\text{AR2/SIM-STARTS}}$



It is interesting to observe the dynamics of forecasts for the SIM-STARTS in the pre-crisis, crisis and post-crisis sub-samples altogether spanning the period of 2002–2011. In the pre-crisis period, the CSSFED displayed in Figure B.3(b) fluctuates around some level, indicating that the forecasting performance of the SIM-STARTS model is comparable with that of the benchmark model. The situation, however, drastically changes during the Great Recession period when suddenly the STARTS variable gains forecasting power. Nevertheless, in the post-crisis period this forecasting power evaporates as rapidly as it appeared during the crisis. In fact, as shown in Figure B.2(b), the SIM-STARTS forecasts in this period severely underestimate the GDP growth rate at FO4.

B.2 Density forecasts

The forecasting performance of the SIMs is summarised in Table B.2 in terms of the ALSs and ALSDs against the benchmark AR(2)-SV model for the full sample as well as for its two sub-samples. Comparing these outcomes with the results reported in Sub-section B.1 for point forecasts, a number of striking similarities can be noticed. First, we also observe heterogeneous results for different indicator models. The three best models (SIM-ISM, SIM-EMPLOY, SIM-ORDERS) that produce most accurate point forecasts also produce most accurate density forecasts in comparison with the benchmark model as well as the rest of the SIMs. Second, the asymmetric forecasting performance during expansions and recessions is also present for density forecasts. Third, in terms of relative forecasting performance with respect to the benchmark model the differences are much more pronounced during recessions than expansions. The dynamics of the forecasting performance of selected models relative to the benchmark model over time can be visually assessed in Figure B.4. Fourth, the backcasts made at the forecast origin FO4 produce the most accurate density forecasts compared with those made at the earlier forecast horizons.

A summary over the forecasting performance of the SIMs over forecast origins is presented in Figure B.5. The left panel contains the results for the full sample. Here the standard results are present: accuracy of density forecasts seems to improve as the forecast origins advance in time. However, as it can be seen from the middle and right panels, the magnitude of improvements in density forecasts over the benchmark model is much more pronounced for recessions than expansions.

Table B.2
Density forecast accuracy, SIMs

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$ALSD_{\bullet/AR(2)\text{-}SV}$*			
Full sample								
AR(2)-SV	2.253	2.166	2.159	2.182				
SIM-ISM-SV	2.249	2.130	2.072	2.024	–0.004	–0.036	–0.087	–0.158
SIM-EMPLOY-SV	2.233	2.148	2.094	2.068	–0.020	–0.018	–0.066	–0.114
SIM-SUPDEL-SV	2.251	2.175	2.154	2.153	–0.002	0.010	–0.006	–0.029
SIM-ORDERS-SV	2.221	2.112	2.045	2.004	–0.032	–0.053	–0.114	–0.178
SIM-HOURS-SV	2.262	2.184	2.175	2.157	0.009	0.018	0.015	–0.025
SIM-SP500-SV	2.247	2.182	2.158	2.156	–0.007	0.016	–0.001	–0.026
SIM-TBILL-SV	2.260	2.181	2.183	2.164	0.007	0.015	0.023	–0.019
SIM-TBOND-SV	2.259	2.180	2.170	2.159	0.006	0.015	0.011	–0.023
SIM-CLAIMS-SV	2.272		2.150	2.112	0.019		–0.009	–0.070
SIM-RSALES-SV	2.268		2.170	2.160	0.015		0.010	–0.023
SIM-IP-SV	2.252		2.137	2.115	–0.001		–0.023	–0.068
SIM-STARTS-SV	2.242		2.124	2.126	–0.011		–0.036	–0.056
Boom sample								
AR(2)-SV	2.092	2.043	2.044	2.059				
SIM-ISM-SV	2.087	2.023	1.990	1.959	–0.005	–0.020	–0.054	–0.100
SIM-EMPLOY-SV	2.083	2.032	2.009	2.014	–0.009	–0.011	–0.035	–0.045
SIM-SUPDEL-SV	2.090	2.047	2.034	2.032	–0.002	0.005	–0.010	–0.027
SIM-ORDERS–SV	2.069	2.005	1.969	1.948	–0.022	–0.038	–0.075	–0.111
SIM-HOURS-SV	2.101	2.062	2.063	2.057	0.009	0.019	0.019	–0.002
SIM-SP500-SV	2.082	2.053	2.044	2.043	–0.010	0.010	–0.000	–0.015
SIM-TBILL-SV	2.079	2.036	2.047	2.039	–0.013	–0.007	0.003	–0.020
SIM-TBOND-SV	2.086	2.050	2.041	2.039	–0.006	0.007	–0.003	–0.019
SIM-CLAIMS-SV	2.095		2.034	2.021	0.003		–0.010	–0.037
SIM-RSALES-SV	2.095		2.038	2.038	0.003		–0.006	–0.021
SIM-IP-SV	2.084		2.020	2.012	–0.008		–0.024	–0.047
SIM-STARTS-SV	2.082		2.042	2.058	–0.009		–0.002	–0.001
Bust sample								
AR(2)-SV	3.324	2.982	2.926	3.004				
SIM-ISM-SV	3.324	2.844	2.617	2.458	0.000	–0.138	–0.309	–0.547
SIM-EMPLOY-SV	3.232	2.918	2.657	2.432	–0.092	–0.063	–0.269	–0.572
SIM-SUPDEL-SV	3.320	3.025	2.951	2.960	–0.004	0.043	0.025	–0.044
SIM-ORDERS-SV	3.231	2.825	2.551	2.378	–0.093	–0.156	–0.375	–0.626
SIM-HOURS-SV	3.333	2.991	2.914	2.821	0.009	0.009	–0.012	–0.183
SIM-SP500-SV	3.340	3.039	2.916	2.908	0.016	0.058	–0.010	–0.096
SIM-TBILL-SV	3.466	3.144	3.083	2.996	0.141	0.163	0.157	–0.008
SIM-TBOND-SV	3.411	3.045	3.026	2.956	0.087	0.064	0.100	–0.048
SIM-CLAIMS-SV	3.446		2.918	2.714	0.122		–0.008	–0.290
SIM-RSALES-SV	3.416		3.044	2.967	0.091		0.118	–0.037
SIM-IP-SV	3.372		2.914	2.797	0.048		–0.012	–0.207
SIM-STARTS-SV	3.302		2.664	2.577	–0.022		–0.262	–0.427

* The symbol • in $ALSD_{\bullet/AR(2)\text{-}SV}$ denotes the name of a competing SIM.

Figure B.4
ALS summary, SIMs

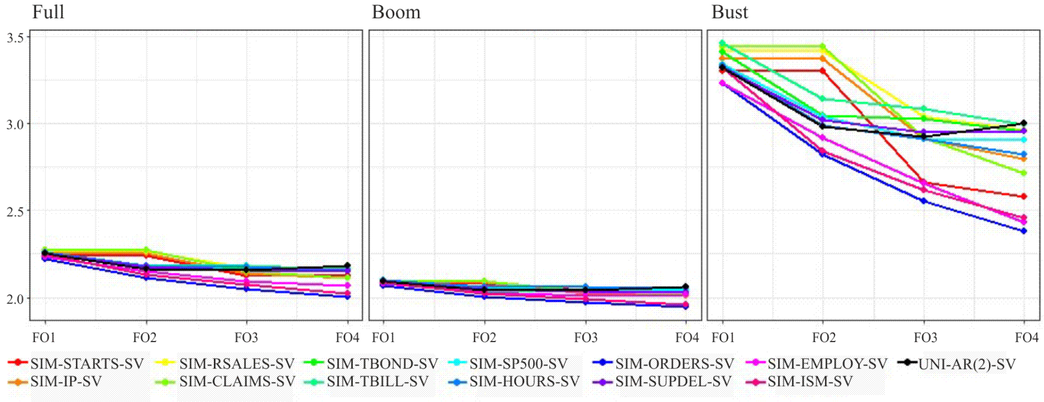
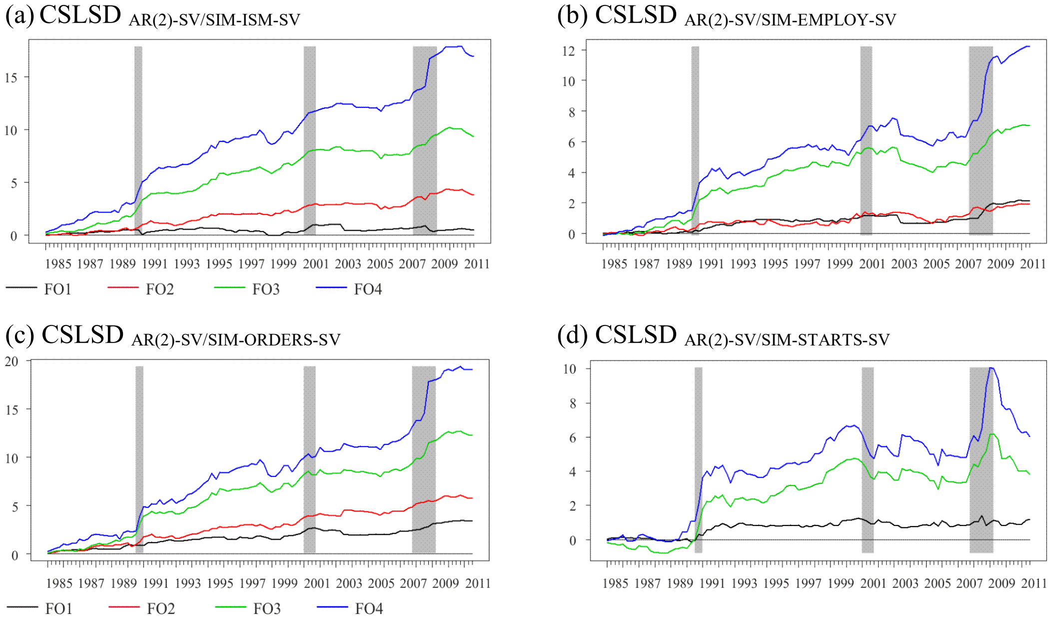


Figure B.5
Density forecast accuracy, SIMs



All in all, consistent with the results reported above for point forecasts care should be taken when assessing density forecasting performance over longer periods of time that span one or several business cycles. Ignoring the differential performance of the models during business cycle phases is likely to lead to distorted evaluation of model predictive ability in normal and crisis times.

C. Combinations of SIMs

C.1 Point forecasts

The results of the forecasting exercise using combinations of SIMs based on their point forecasting performance is summarised in Table C.1 in terms of RMSFE and RRMSFE. In general, the conclusions drawn from Table B.1 apply also for their combinations. First, there is asymmetry in the forecasting performance across business cycle phases which again biases conclusions on the forecasting performance of model combinations reported for the whole forecasting sample ignoring booms and busts. Second, gains in forecasting accuracy over the benchmark model are mainly brought about by those observations during recessions. It is worthwhile noticing that in expansions the forecasting performance of model combinations is comparable to that of the most accurate SIMs (see Table B.1). However, this is not the case during recessions; the forecasting performance of model combinations is somewhat lower than that of the best SIM. Out of model combinations the most accurate forecasts during recessions were produced by combinations based on recursive weighting with a rather heavy discounting, $\delta = 0.30$. It is worthwhile noticing that during recessions the accuracy of forecasts based on equal weighting were found to be inferior to that based on recursive weighting.

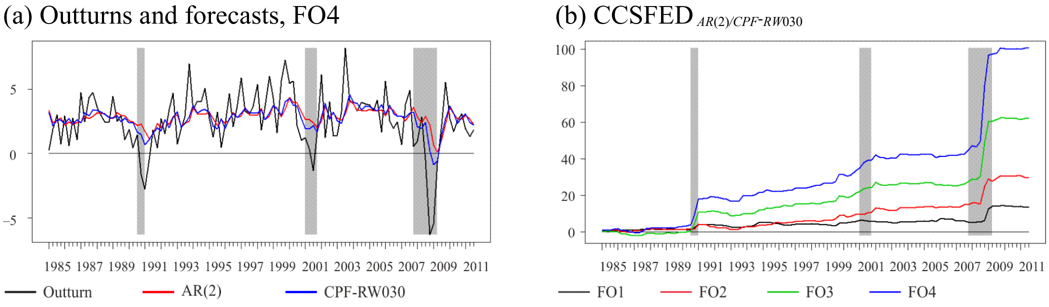
Table C.1
Point forecast accuracy, point forecast combination of SIMs

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				RRMSFE$_{RW/AR(2)}$*			
Full sample								
AR(2)	2.237	2.102	2.081	2.074				
CPF-EW	2.210	2.041	1.965	1.886	–0.012	–0.029	–0.056	–0.091
CPF-RW100	2.209	2.037	1.955	1.871	–0.012	–0.031	–0.060	–0.098
CPF-RW090	2.208	2.036	1.949	1.863	–0.013	–0.032	–0.063	–0.102
CPF-RW030	2.207	2.035	1.936	1.833	–0.013	–0.032	–0.070	–0.116
Boom sample								
AR(2)	1.710	1.705	1.692	1.691				
CPF-EW	1.692	1.657	1.633	1.590	–0.010	–0.029	–0.035	–0.059
CPF-RW100	1.692	1.657	1.633	1.588	–0.010	–0.028	–0.035	–0.061
CPF-RW090	1.693	1.657	1.638	1.596	–0.010	–0.028	–0.032	–0.056
CPF-RW030	1.699	1.664	1.646	1.608	–0.006	–0.024	–0.027	–0.049
Bust sample								
AR(2)	4.339	3.802	3.751	3.726				
CPF-EW	4.279	3.687	3.434	3.222	–0.014	–0.030	–0.085	–0.135
CPF-RW100	4.277	3.671	3.392	3.160	–0.014	–0.035	–0.096	–0.152
CPF-RW090	4.268	3.665	3.351	3.097	–0.016	–0.036	–0.107	–0.169
CPF-RW030	4.250	3.640	3.260	2.920	–0.020	–0.043	–0.131	–0.216

* The symbol • in RRMSFE$_{\bullet/AR(2)}$ denotes the name of a competing model.

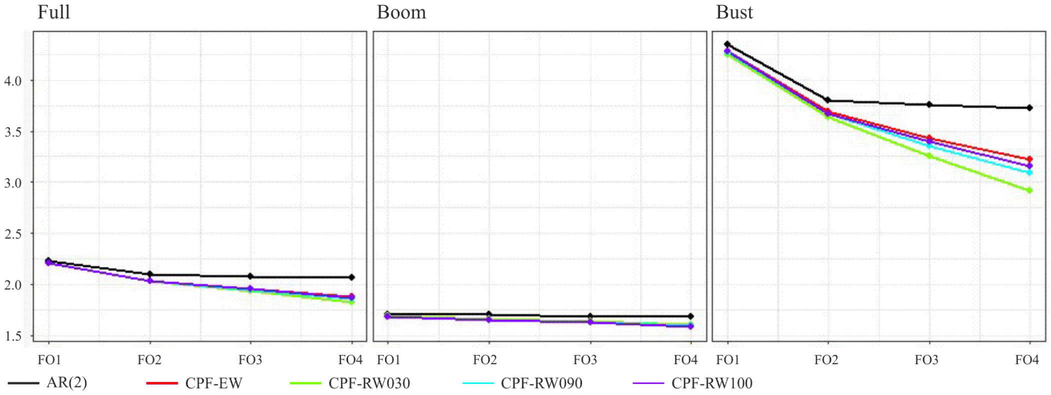
The actual and forecast values for the best forecasting model combination approach are shown in Figure C.1(a) at FO4. The corresponding CSSFEDs for all forecast origins FO1–FO4 are shown in Figure C.1(b). The shape of CSSFED for model combinations is very similar to that of the best model among the SIMs, SIM-ORDERS (see Figure B.3(b)), i.e. jumps occur precisely during recessions.

Figure C.1
Assessment of point forecast accuracy, AR(2) and CPF-RW030



In the left and right panels of Figure C.2 we show the evolution of the RMSFEs by the forecast origin when these are evaluations for the whole sample and when the evaluation is done separately for expansions (the middle panel) and recessions (the right panel). The following comments can be made based on the figure.

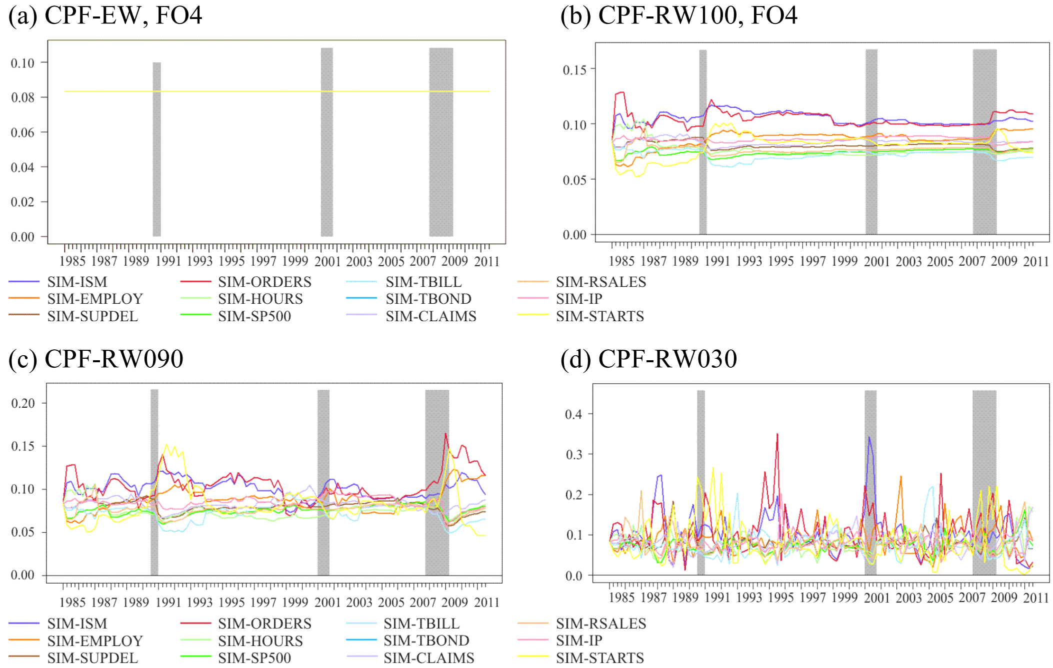
Figure C.2
RMSFE summary, point forecast combinations



First, the model combinations display very little reduction in the reported RMSFEs during expansions. Second, consistent with the results reported for the individual models most of the evidence bringing down average squared forecast errors accrues during resessions. Third, as mentioned above, among all model combinations the one with recursive weights with heavy discounting of the past (CPF-RW030) performs slightly better than the rest of the weighting schemes. Fourth, none of the model combinations is able to improve upon the forecasting accuracy of the best single-indicator model SIM-ORDERS for recessions.

Weights of the model combinations for the latest forecast origin FO4 are shown in Figure C.3. Comparing Figures C.3(a) and C.3(b) one can conclude that going from equal weighting to recursive one without discounting brings about very little: all the models are still assigned almost equal weights. However, by changing the discounting parameter from $\delta = 1$ to $\delta = 0.9$ and further to $\delta = 0.3$ volatility in the weight magnitude increases though none of the models gets assigned a weight approaching the value of 100%.

Figure C.3
Point forecast combinations: weights at FO4, without SV



C.2 Density forecasts

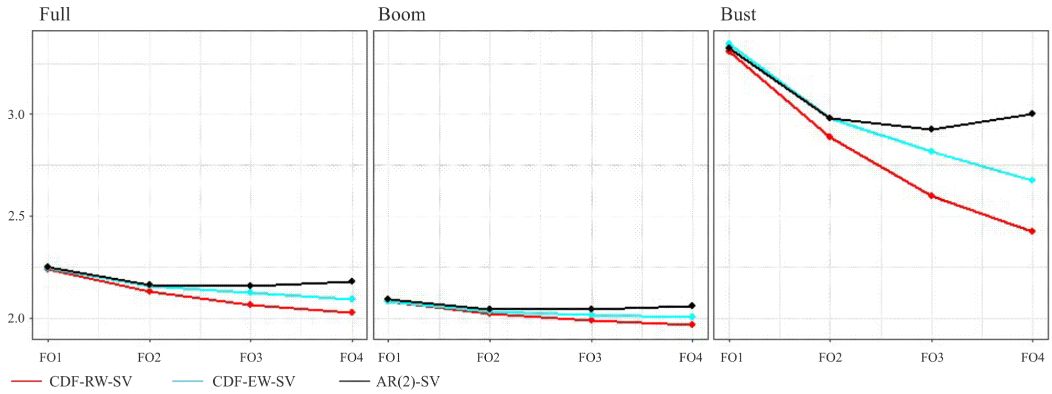
The results of the forecasting exercise using combinations of SIMs based on their density forecasting performance are summarised in Table C.2 in terms of the ALS and ALSD. In Figure C.4 the evolution of the ALS is shown across forecast origins for the whole sample as well as its expansionary and recessionary sub-periods. In general, the conclusions drawn for individual indicator-augmented models apply also for their combinations. First, there is asymmetry in the forecasting performance across business cycle phases. Second, the largest gains in forecasting accuracy over the benchmark model are brought about by those observations during recessions. In addition, as the forecast origin moves forward the increase in the forecasting accuracy is much more pronounced during recessions than expansions; compare the middle and right panels of Figure C.4. Third, the combination based on recursive weighting produces more accurate density forecasts than the scheme based on equal weighting. This holds both for recessions and expansions.

Table C.2
Density forecast accuracy, density forecast combination of SIMs

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				ALSD$_{\bullet/AR(2)\text{-}SV}$*			
Full sample								
AR(2)-SV	2.253	2.166	2.159	2.182				
CDF-EW-SV	2.247	2.157	2.123	2.093	–0.007	–0.009	–0.037	–0.090
CDF-RW-SV	2.241	2.133	2.068	2.026	–0.012	–0.033	–0.091	–0.156
Boom sample								
AR(2)-SV	2.092	2.043	2.044	2.059				
CDF-EW-SV	2.081	2.033	2.018	2.005	–0.010	–0.010	–0.026	–0.054
CDF-RW-SV	2.081	2.019	1.988	1.966	–0.011	–0.023	–0.056	–0.092
Bust sample								
AR(2)-SV	3.324	2.982	2.926	3.004				
CDF-EW-SV	3.343	2.982	2.819	2.677	0.019	–0.000	–0.107	–0.327
CDF-RW-SV	3.308	2.886	2.598	2.423	–0.016	–0.096	–0.328	–0.581

* The symbol • in ALSD$_{\bullet/AR(2)\text{-}SV}$ denotes name of a competing model.

Figure C.4
ALS summary, density forecast combinations



The performance of the density combinations relative to the benchmark model can be assessed by examining the CSLSDs shown in Figure C.5. The largest gains in forecasting accuracy are observed during recessions, especially at the forecasting origin FO4. At the earlier forecasting origins FO1–FO2 the relative forecasting gains are of a somewhat smaller magnitude. It is interesting to note that in the period between the first and second recessions in our sample the model combinations produced more or less steady gains in forecasting accuracy over the benchmark model at FO3–FO4 as evident from the upwards trending CSLSDs.

The weights attached to every SIM are shown in Figure C.6 for all forecast origins. A close examination explains why the description of the forecasting performance of model combinations is very similar to that of individual models. At FO1–FO3 there is only one model (SIM-ORDERS) that dominates the combination.

There are two models (SIM-ORDERS and SIM-ISM) that essentially dominate the combination at FO4.

Earlier in the sample a larger weight is attached to the latter model, whereas during the period of the Great Recession their ranking is reversed with the former model gaining in importance. The weighting scheme based on equation (12) for combinations of density forecasts is much more aggressive than that based on the discounted MSFE in equation (11) for combinations of point forecasts (see Figure C.3 for comparison).

Figure C.5
Density forecast accuracy, density forecast combinations of SIMs

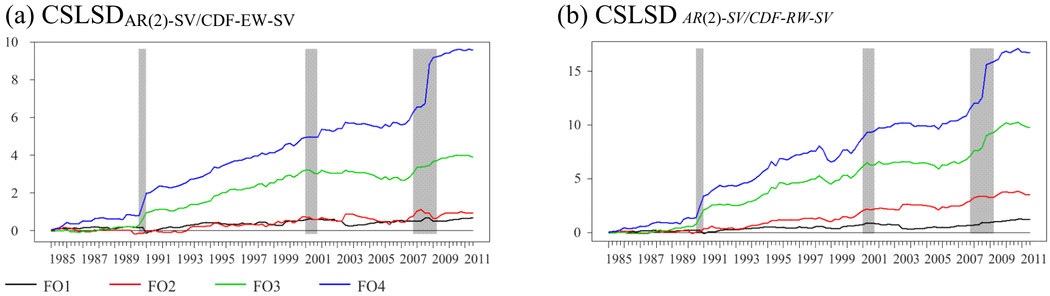
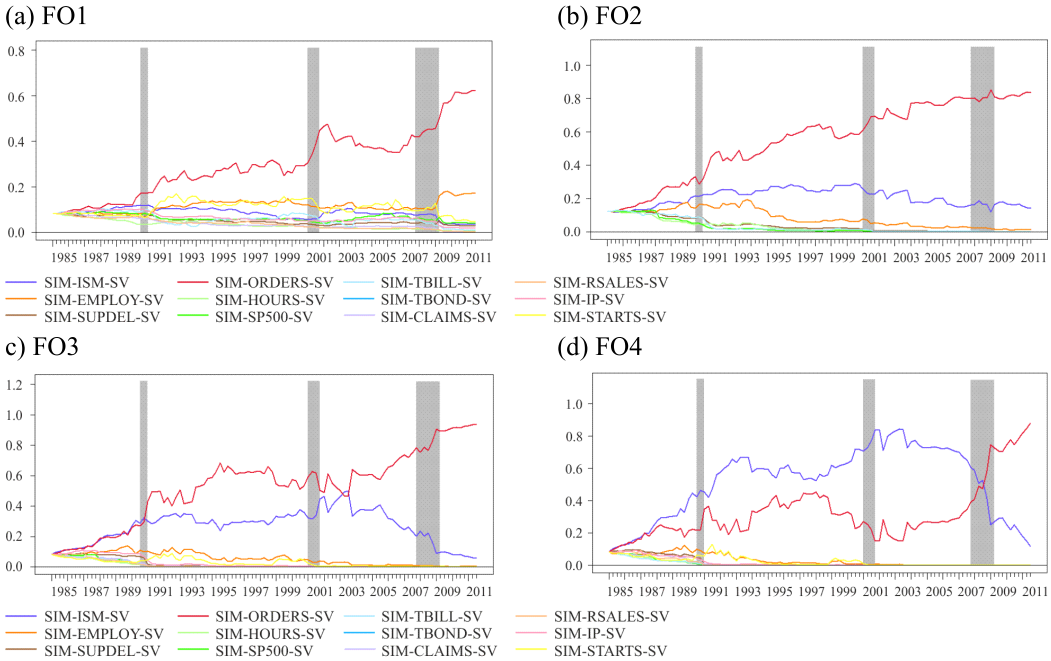


Figure C.6
Density forecast combinations, recursive weights



D. Impact of stochastic volatility on forecast accuracy

D.1 Point forecasts

In this section we will briefly discuss the influence of introducing stochastic volatility into the models evaluated in terms of point forecast accuracy in the Appendix. The relative RMSFEs computed for each model with stochastic and constant volatility are presented in Table D.1. In general, we find no systematic evidence that the addition of stochastic volatility improves accuracy of point forecasts. It is interesting that this finding holds both for the whole forecast evaluation sample and the two boom/bust sub-periods. The magnitude of the effect in most cases comprises a couple of percentage points in either direction depending on a model type and forecast origin. This finding is consistent with the results reported in Table 2 of Carriero et al. (2015,

p. 849) that also compares forecasting accuracy of models with constant and stochastic volatility for the whole forecast evaluation sample.

Table D.1
Impact of stochastic volatility on point forecast accuracy

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	Full sample				**Boom sample**				**Bust sample**			
AR(2)	0.010	0.002	0.005	0.003	0.015	0.011	0.015	0.008	0.004	–0.010	–0.009	–0.004
RW	0.009	0.003	0.017	0.001	0.003	–0.006	0.011	–0.005	0.015	0.010	0.022	0.007
SIM-ISM	0.008	0.010	0.007	–0.001	0.007	0.011	–0.005	–0.005	0.008	0.010	0.031	0.009
SIM-EMPLOY	0.008	0.009	0.008	0.010	0.007	0.001	–0.001	0.006	0.008	0.022	0.027	0.023
SIM-SUPDEL	0.008	0.005	–0.009	–0.001	0.013	0.005	–0.002	0.006	0.002	0.005	–0.018	–0.010
SIM-ORDERS	0.008	0.009	0.019	0.010	0.017	0.002	0.002	–0.008	–0.004	0.019	0.057	0.063
SIM-HOURS	–0.008	–0.003	–0.004	–0.013	–0.006	–0.001	–0.002	–0.013	–0.011	–0.006	–0.008	–0.014
SIM-SP500	0.001	0.006	0.018	0.008	0.003	0.011	0.014	0.013	–0.001	0.000	0.024	0.001
SIM-TBILL	–0.017	–0.028	–0.019	–0.024	0.026	0.010	0.021	0.014	–0.054	–0.068	–0.063	–0.066
SIM-TBOND	–0.015	–0.011	–0.010	–0.013	0.019	0.023	0.029	0.022	–0.045	–0.049	–0.057	–0.056
SIM-CLAIMS	0.004		0.005	0.000	0.011		0.007	0.011	–0.003		0.001	–0.020
SIM-RSALES	0.001		–0.004	0.002	0.014		0.001	0.003	–0.013		–0.010	0.002
SIM-IP	0.002		0.005	0.006	0.010		0.009	0.007	–0.006		–0.001	0.005
SIM-STARTS	0.007		–0.012	–0.039	–0.016		–0.039	–0.076	0.036		0.057	0.092
CPF-EW	0.004	0.004	0.010	0.009	0.015	0.015	0.014	0.013	–0.007	–0.010	0.003	0.003
CPF-RW100	0.004	0.005	0.011	0.010	0.015	0.015	0.014	0.012	–0.007	–0.008	0.007	0.008
CPF-RW090	0.005	0.006	0.013	0.012	0.015	0.015	0.013	0.010	–0.005	–0.007	0.013	0.016
CPF-RW030	0.005	0.006	0.017	0.020	0.013	0.012	0.011	0.009	–0.002	–0.003	0.026	0.042

The table entries are the values of RRMSFE computed for each model with stochastic and constant volatility. The symbol • in $RRMSFE_{\bullet-SV/\bullet}$ denotes the name of the respective model.

D.2 Density forecasts

The conclusion that there is no systematic effect on accuracy of point forecasts from adding stochastic volatility to the forecasting models was stated in Section D.1. In this section we intend to investigate the same question for density forecasts. As above, we address this question for the full sample and its boom and bust sub-samples.

Table D.2
Impact of stochastic volatility on density forecast accuracy

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	Full sample				**Boom sample**				**Bust sample**			
AR(2)	–0.212	–0.245	–0.249	–0.231	–0.297	–0.311	–0.309	–0.300	0.356	0.199	0.154	0.227
RW	–0.182	–0.212	–0.228	–0.231	–0.293	–0.300	–0.299	–0.299	0.556	0.375	0.240	0.225
SIM-ISM	–0.193	–0.231	–0.238	–0.244	–0.285	–0.292	–0.286	–0.283	0.418	0.175	0.077	0.013
SIM-EMPLOY	–0.211	–0.200	–0.211	–0.206	–0.294	–0.267	–0.264	–0.239	0.335	0.250	0.142	0.012
SIM-SUPDEL	–0.209	–0.244	–0.255	–0.241	–0.292	–0.316	–0.323	–0.310	0.343	0.235	0.199	0.214
SIM-ORDERS	–0.205	–0.227	–0.210	–0.209	–0.290	–0.292	–0.260	–0.247	0.353	0.204	0.120	0.041
SIM-HOURS	–0.202	–0.244	–0.250	–0.262	–0.297	–0.315	–0.318	–0.319	0.425	0.225	0.197	0.120
SIM-SP500	–0.211	–0.226	–0.251	–0.243	–0.303	–0.306	–0.317	–0.309	0.397	0.306	0.188	0.199
SIM-TBILL	–0.202	–0.242	–0.230	–0.253	–0.281	–0.308	–0.295	–0.303	0.322	0.193	0.203	0.082
SIM-TBOND	–0.209	–0.245	–0.246	–0.244	–0.293	–0.309	–0.313	–0.310	0.349	0.183	0.198	0.189
SIM-CLAIMS	–0.187		–0.246	–0.263	–0.287		–0.315	–0.314	0.482		0.209	0.079
SIM-RSALES	–0.194		–0.238	–0.243	–0.287		–0.314	–0.313	0.423		0.263	0.216
SIM-IP	–0.194		–0.224	–0.205	–0.290		–0.273	–0.255	0.445		0.102	0.126
SIM-STARTS	–0.190		–0.234	–0.222	–0.292		–0.290	–0.277	0.486		0.142	0.139
CDF-EW	–0.199	–0.230	–0.238	–0.245	–0.289	–0.300	–0.299	–0.296	0.397	0.235	0.168	0.097
CDF-RW	–0.200	–0.219	–0.206	–0.206	–0.287	–0.286	–0.260	–0.248	0.379	0.231	0.147	0.070

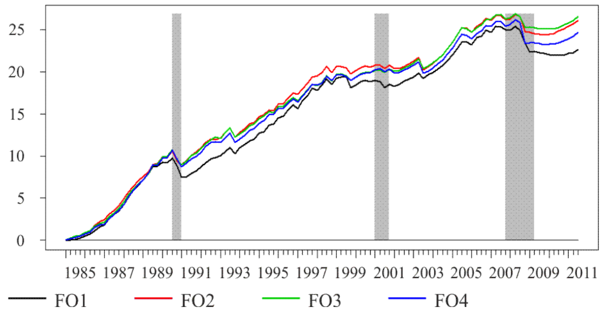
The table entries are the values of ALSD computed for each model with stochastic and constant volatility. The symbol • in ALSD$_{\bullet-SV/\bullet}$ denotes the name of the respective model.

The effect of the stochastic volatility on the density forecast accuracy is reported in Table D.2. Entries in the table are differences in ALS reported between models with stochastic and constant volatility in the residual error term. Negative entries imply that on average density forecast accuracy of the models with stochastic volatility is higher, postive entries indicate the opposite. As seen, for the full sample as well as for expansions the models with stochastic volatility produce lower ALS than their counterparts with constant volatility. This is in line with results reported in Carriero et al. (2015). However, for recessions we have a different situation. For the benchmark models, the SIMs and SIM combinations, there are positive entries at all forecast origins FO1–FO4 indicating that during recessions models with stochastic volatility produced on average less accurate density forecasts.

In tracking the reason for this at the first glance surprising result it is instructive to inspect plots of CSLSD for the models with stochastic and constant volatility. These are displayed for the benchmark AR(2) model in Figure D.1. In general, upwards trending behaviour of the CSLSDs indicates more or less steadily gains in density forecast accuracy of the model with stochastic volatility over that with constant volatility. However, there are several observations when the latter models produced more accurate density forecasts. Namely, these observations belong to the recession period in the early 1990s and the Great Recession.

This finding reinforces the observation made in the main text for the density forecasts of the multiple-indicator models. During recessions when actual outcomes of the predicted variable lies farther out in the tails of the variable distribution the fatter tails of the models with constant volatility may better capture uncertainty surrounding nowcasts. This observation at best can be seen and explained when a comparison of the relative forecasting performance of the competing models is done by using recursive forecast evaluation metrics. As demonstrated in Carriero et al. (2015), when relying on the measures based on the averages computed for the whole forecast evaluation sample, this peculiar finding is likely to be left unnoticed by the researcher.

Figure D.1
Impact of stochastic volatility on density forecast accuracy: CSLSD$_{AR(2)\text{-}SV/AR(2)}$



Summary

The results presented in the Appendix confirm the findings reported in the main text. The asymmetry in the absolute and relative measures of the accuracy of point and density forecasts across the business cycle phases can be easily detected for these additional models considered. Our results indicate that the data-pooling approach of Carriero et al. (2015) produces more accurate point and density forecasts during recessions than the model-pooling approach of Mazzi et al. (2014). During expansions their performance is more or less similar. The impact of stochastic volatility on point forecast accuracy is unsystematic and of a relatively minor magnitude. However, the impact of stochastic volatility on density forecasts is positive during expansions, but during recessions the impact sign is reverted, i.e. models with constant volatility produce more accurate density forecasts than their counterparts with stochastic volatility.

BIBLIOGRAPHY

CARRIERO, Andrea, CLARK, Todd E., MARCELLINO, Massimiliano (2013). Real-Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. CEPR Discussion Paper, No. 9312. 38 p.

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.

CHAN, Joshua C. C. (2017). The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling. *Journal of Business & Economic Statistics*, vol. 35, issue 1, pp. 17–28.

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

CLARK, Todd E. (2011). Real-Time Density Forecasts from Bayesian Vector Autoregressions with Stochastic Volatility. *Journal of Business & Economic Statistics*, vol. 29, issue 3, pp. 327–341.

CROUSHORE, Dean, STARK, Tom (2001). A Real-Time Data Set for Macroeconomists. *Journal of Econometrics*, November 2001, vol. 105, issue 1, pp. 111–130.

CROUSHORE, Dean, STARK, Tom (2003). A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter? *The Review of Economics and Statistics*, vol. 85, issue 3, August 2003, pp. 605–617.

EDGE, Rochelle M., KILEY, Michael T., LAFORTE, Jean-Philippe (2010). A Comparison of Forecast Performance between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model. *Journal of Applied Econometrics*, vol. 25, No. 4, June–July 2010, pp. 720–754.

FORONI, Claudia, MARCELLINO, Massimiliano (2013). *A Survey of Econometric Methods for Mixed-Frequency Data.* Norges Bank Working Paper, No. 2013/06. 42 p.

FORONI, Claudia, MARCELLINO, Massimiliano, SCHUMACHER, Christian (2015). Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 178, issue 1, pp. 57–82.

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.

GHYSELS, Eric, SANTA-CLARA, Pedro, VALKANOV, Rossen (2004). *The MIDAS Touch: Mixed Data Sampling Regression Models.* CIRANO Working Papers, May 2004, No. 2004s-20. 36 p.

GHYSELS, Eric, SINKO, Arthur, VALKANOV, Rossen (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, February 2007, vol. 26, issue 1, pp. 53–90.

GIANNONE, Domenico, MONTI, Francesca, REICHLIN, Lucrezia (2016). Exploiting the Monthly Data Flow in Structural Forecasting. *Journal of Monetary Economics*, vol. 84, issue C, pp. 201–215.

GNEITING, Tilmann, KATZFUSS, Matthias (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, vol. 1, pp. 125–151.

GUÉRIN, Pierre, MARCELLINO, Massimiliano (2013). Markov-Switching MIDAS Models. *Journal of Business & Economic Statistics*, vol. 31, issue 1, pp. 45–56.

KIM, Hyun Hak, SWANSON, Norman R. (2018). Mining Big Data Using Parsimonious Factor, Machine Learning, Variable Selection and Shrinkage Methods. *International Journal of Forecasting*, vol. 34, issue 2, pp. 339–354.

MARCELLINO, Massimiliano, PORQUEDDU, Mario, VENDITTI, Fabrizio (2016). Short-Term GDP Forecasting with a Mixed-Frequency Dynamic Factor Model with Stochastic Volatility. *Journal of Business & Economic Statistics*, January 2016, vol. 34, issue 1, pp. 118–127.

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.

MAZZI, Gian Luigi, MITCHELL, James, MONTANA, Gaetana (2014). Density Nowcasts and Model Combination: Nowcasting Euro-Area GDP Growth over the 2008–09 Recession. *Oxford Bulletin of Economics and Statistics*, Department of Economics, University of Oxford, vol. 76, issue 2, pp. 233–256.

McCRACKEN, Michael W., OWYANG, Michael T., SEKHPOSYAN, Tatevik (2015). *Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR.* Working Paper No. 2015-030A, Federal Reserve Bank of St. Louis, October 2015. 34 p.

MIKOSCH, Heiner F., NEUWIRTH, Stefan Gerald (2015). *Real-Time Forecasting with a MIDAS VAR.* KOF Working Papers, No. 15-377, KOF Swiss Economic Institute, ETH Zurich, 33 p.

SCHORFHEIDE, Frank, SONG, Dongho (2015). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*, vol. 33, issue 3, pp. 366–380.

SILIVERSTOVS, Boriss (2017). Short-term Forecasting with Mixed-Frequency Data: A MIDASSO Approach. *Applied Economics*, vol. 49, issue 13, pp. 1326–1343.

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

ZHANG, Bo, CHAN, Joshua, C. C., CROSS, Jamie L. (2018). *Stochastic Volatility Models with ARMA Innovations: An Application to G7 Inflation Forecasts.* CAMA Working Paper No. 32/2018, June 2018. 32 p.