Assessing Nowcast Accuracy of US GDP Growth in Real Time: The Role of Booms and Busts[*]

Boriss Siliverstovs
Latvijas Banka,
Kr. Valdemara street 2a,
LV-1050, Riga, Latvia
boriss.siliverstovs@bank.lv

December 28, 2017

Abstract

In this paper we re-assess 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 the US macroeconomic data. Following Chauvet and Potter (2013), we evaluate the forecasting accuracy of the suggested model and the univariate benchmark models separately for expansions and recessions as defined by the NBER business cycle chronology, rather than relying on the comparison of forecast accuracy over the whole forecast sample spanning 1985Q1—2011Q3. We find that most of the evidence favouring more sophisticated model over its simple benchmarks is due to relatively few observations during the recessions, especially, those during the Great Recession. In contrast, during expansions the gains in forecasting accuracy over the benchmark models are very modest at best. This means that the relative forecasting performance of the models in question varies with business cycle phases. Ignoring this fact results in a distorted picture: the relative performance of a more sophisticated model in comparison with a naive benchmark model during expansions tends to be overstated and, consequently, during recessions—understated.

Keywords: nowcasting, mixed-frequency data, real-time data, business cycle
JEL code: C22, C53

[1]The paper benefited from comments made by the seminar participants at the Department of Economics and Finance, Tallinn University of Technology in Estonia and the 18th IWH-CIREQ-GW Macroeconometric workshop "Mixed frequency data in macroeconomics and finance" in Halle, Germany. All computations were performed in R and RcppArmadillo. Any remaining errors are the sole responsibility of the author. The views expressed are mine and do not necessarily reflect those of the *Latvijas Banka*.

1 Introduction

In a recent contribution to the *Handbook of Forecasting* Chauvet and Potter (2013) provided a comprehensive review of forecasting performance of several state-of-the-art econometric models. These models include univariate linear autoregressive model of order two, AR(2), univariate non-linear models that takes into account Cumulative Depth of Recession (CDR) and Markov-switching dynamics (MS), Dynamic Stochastic General Equilibrium (DSGE) model, Bayesian and non-Bayesian Vector AutoRegressive (BVAR and VAR) models as well as 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 of Chauvet and Potter (2013).

Utilising forecasts of those models of the quarterly GDP growth rate in the US over the period spanning almost 20 years (1992Q1—2010Q4) Chauvet and Potter (2013) conducted assessment of forecasting accuracy of these models not only for the full sample but also for its sub-samples of expansions and recessions as defined by the NBER business cycle chronology. The results reported in Chauvet and Potter (2013) can be summarised as follows. First, the absolute size of forecast errors varies with the business cycle phases with those being larger during recessions. This indicates that for the models in question it was more difficult to forecast GDP growth during recessions than expansions. Second, during expansions all models considered produced average forecast accuracy that at best is comparable to one from the AR(2) model, which is a popular choice for a benchmark model in the literature.[1] In a way this findings is very consonant with the results reported in Nelson (1972) about 40 years ago showing that the forecast accuracy of the large and very sophisticated model jointly developed at the Federal Reserve Board, MIT and Penn was generally inferior to out-of-sample predictions from simple-to-run univariate autoregressive models. Third, Chauvet and Potter (2013) conclude that during recessions forecasts from only one econometric model (AR-DFMS) as well as Blue Chip forecasts are more accurate in a statistically meaningful way than those from the benchmark model. The DSGE model produces slightly more accurate forecasts but their accuracy is not statistically different from one of the benchmark model. The (B)VAR-model based forecasts fare uniformly worse than the benchmark model.

Despite a clearly stated outcome, the idea of Chauvet and Potter (2013) of reporting absolute and relative forecasting performance of models in question separately for recessions and expansions so far did not receive wide response in the macroeconomic forecasting literature. Even papers published later than the paper of Chauvet and Potter (2013) overwhelmingly assess forecasting

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

performance for the whole forecast evaluation period or, at best, for the sub-periods of the pre- and post-Great Recession crisis. For example, in Carriero et al. (2015) the results of nowcasting models are reported for the whole period 1985Q1—2011Q3 and for the pre-crisis period 1985Q1—2008Q2. Foroni et al. (2015) use the following forecast evaluation samples: 1985Q1—2011Q1 and 1985Q1—2006Q4. In Giannone et al. (2016) the results are reported for the whole forecast evaluation sample 1995Q1—2014Q2 and the pre- and post-crisis sub-samples, 1995Q1—2007Q4 and 2008Q1—2014Q2. Schorfheide and Song (2015) uses the sample 1997Q2—2009Q4 for evaluation of forecasts of the suggested mixed-frequency BVAR model. Kim and Swanson (2017) also ignore the possibility of heterogeneous forecasting accuracy across business cycle phases when reporting their results for one of the longest forecast evaluation samples 1974M03—2009M05.

By mixing in one way or another periods of recessions and expansions, these studies are likely to conceal differences in forecasting performance across recessions and expansions and therefore present a distorted picture of models' predictive ability. A notable exception from the mentioned studies is Bulligan et al. (2014) where the authors explicitly divide a relatively short total forecast evaluation sample consisting of 13 quarters (2006Q4—2009Q4) into crisis (2008Q2—2009Q2) and non-crisis (2006Q4—2008Q1 and 2009Q3—2009Q4) periods. In doing so, these authors effectively chose the Great Recession period and several quarters shortly before and after the Great Recession. Because of rather few observations involved in the analysis it is not so clear how the conclusion "...that the gains associated with models based on targeted predictors are not driven by few exceptional observations." of Bulligan et al. (2014, p. 203) can be generalised to other occurrences of recessions or longer periods of economic expansions.

In this paper we intend to extend the work of Chauvet and Potter (2013) in several dimensions. First, we extend the analysis of Chauvet and Potter (2013) to a different class of models that take into account intra-quarter flow of information by utilizing monthly economic/financial indicators in a mixed-frequency data framework. Adoption of this mixed-frequency approach to forecasting enables us to track marginal month-to-month improvements in forecasting accuracy evaluated both for total sample as well as individually for its recession and expansion phases as more monthly information gets absorbed into the econometric model. For this purpose, we utilise the Bayesian mixed-frequency model with stochastic volatility proposed in Carriero et al. (2015) for forecasting US GDP growth.

Second, adoption of the Bayesian estimation framework allows us to complement the analysis of Chauvet and Potter (2013) done for point forecasts with a similar analysis for density forecasts. To the best of our knowledge this is the first attempt to dissect density forecasting performance of mixed-frequency models according to the business cycle phases in a fashion similar to how it is done in Chauvet and Potter (2013). Our paper also intends to contribute to still rather small but growing literature that evaluates model forecasting performance not only in terms of point but also density forecasts, see Aastveit et al. (2017) as well as Mazzi et al. (2014), apart from already mentioned Carriero et al. (2015).

Third, since importance of the use of real-time data vintages was emphasized in Chauvet and Potter (2013) we also resort to such a data set. To this end, we utilise the real-time data set compiled by Carriero et al. (2015). The data comprise historical data vintages for each month from January 1985 until October 2011. Such a rich data set allows us to make forecasts for each quarter since 1985Q1 until 2011Q3 using information that was acually available to a forecaster in the past. This means that compared to forecast evaluation sample of Chauvet and Potter (2013) we do our analysis on the sample that is longer by almost eight years. This also makes it possible to evaluate whether conclusions reached in Chauvet and Potter (2013) are robust with respect to the forecast evaluation sample used.

Fourth, we investigate whether the forecasting performance of the multiple-indicator mixed-frequency model proposed in Carriero et al. (2015) can be improved upon by combining multiple single-indicator mixed-frequency models both in terms of point and density forecasts. Though model combinations according to their point forecasting performance has a long tradition in the forecasting literature (Bates and Granger, 1969), model combinations according to their density forecasting performance so far has received much less attention in the literature (see, e.g. Mazzi et al., 2014; Aastveit et al., 2014, 2017).

Last but not least, we would like to emphasise the use of such recursive measures of forecast accuracy as the Cumulated Squared Sum of Forecast Error Difference (CSSFED) of Welch and Goyal (2008) and Cumulated Sum of Logarithmic Score Difference (CSLSD) 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/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 limited in the macroeconomic forecasting literature. Our paper intends to popularise further the use of recusive metrics by providing an illustrative example that in line of the argument of Geweke/Amisano (2010) shows that these recursive measures of forecast accuracy are much more informative about models' comparative forecasting performance than widely used aggregated measures of forecast accuracy computed over the whole sample. Such aggregation of results often mask heterogeneous forecasting performance of the competing models over sub-samples: in our case, over expansions and recessions (see Siliverstovs, 2017a, for an analysis illustrating differences in forecastability of international stock market indices during recessions and expansions).

Our main findigs can be summarised as follows. First, compared with the results of Chauvet and Potter (2013) we also document asymmetric forecasting ability during expansions and recessions. But in our modelling setup this asymmetry typically arises in models that are based on partial information like univariate autoregressive models as well as models augmented with only one economic/financial indicator (labelled as single-indicator models in sequel) and their combinations. For models which are based on sufficiently large information sets and that use this information efficiently, e.g. by data pooling into one model, this asymmetry is only noticable at longer forecast horizons but it eventually vanishes as more monthly information is incorporated into the forecasting model. Remarkably, this conclusion can be drawn not only for point but also for density forecasts.

Second, even though differences in the forecasting accuracy during expansions/recessions eventually disappear for selected models, the differences in the forecasting performance relative to the benchmark model do stay there. Consistent with results of Chauvet and Potter (2013) we also find that during expansions the benchmark univariate autoregressive models or, for that matter, a historical mean model do produce forecasting accuracy that is comparable to one from more sophisticated models. It is only during recessions when these models that use additional information from monthly indicators are able to bring about substantial gains in the forecast accuracy relative to the benchmark models. This asymmetry in the relative forecasting performance across the business cycle phases has strong implications for a message delivered by those studies that ignore it. When comparing the results of a forecasting exercise with naive benchmark models these studies tend to overstate the predictive ability of their preferred and usually more complicated models in expansions and, consequently, to understate it during recessions. In doing so, these studies deliver a biased assessment of model forecasting accuracy to business analysts, policy-makers or any other sides interested in their forecasts.

Third, in the context of the question of whether to pool data or models we obtain a convincing evidence that at least for the data and models at hand the former approach provides more accurate point as well as density forecasts during recessions. Compared to the multiple-indicator models of Carriero et al. (2015) the single-indicator models and thence their combinations (severily) underestimate the depth of the Great Recession, for example. Whereas the models of Carriero et al. (2015)

successfully demonstrate their ability to do so whenever there is sufficient information available.

All in all, our findings in the reinforcement of similar findins of Chauvet and Potter (2013) do have important implications when reporting results of forecasting competitions over time periods characterised by a expansion/recession dichotomy. We suggest separate reporting of the corresponding results over recessions and expansions isolating the effect of most influential observations (typically occurring during recessions) on the measures of the forecast accuracy computed over the whole forecast evaluation sample.

The rest of the paper is structured as follows. The data used are presented in Section 2. The econometric model of Carriero et al. (2015) is described in Section 3. The metrics used in evaluation of point and density forecasts is presented in Section 4. The results for point and density forecasts are discussed in Sections 5 and 6, respectively. The final section concludes.

2 Data

The data used in this paper are real-time data vintages collected in Carriero et al. (2015). These vintages comprise quarterly GDP data as well as twelve monthly indicators that are widely considered as ones of the most important economic/financial indicators for assessing economic conditions in the US. The data are presented in Table 1. The data represent business tendency surveys (ISM, SUPDEL, ORDERS), labour market conditions (EMPLOY, CLAIMS), production (IP, HOURS) as well consumption (RSALES) sides of the economy, housing (HS) and financial (SP500, TBILL, TBOND) markets.

Carriero et al. (2015) produce forecasts of GDP growth for each quarter since 1985Q1 until 2011Q3. For each quarter t in question the following four forecast origins (FO1, FO2, FO3, FO4) are set when the respective forecasts are made: FO1 — the end of the first week of the first month in quarter t, FO2 and FO3 — the end of the first week of the second and third months in quarter t, respectively, and FO4 — 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 as *nowcasts*, i.e. those made for now or the current quarter, and forecasts made at FO4 — *backcasts*, i.e. those that are backwards-looking or that are made (shortly) after the end of the targeted quarter.

The last column in Table 1 indicates whether observations for the previous month are available during the first or second week of the current month. These differences in release timing of monthly indicators have important implications for their availability at a given forecast origin. For example, at FO1 in a given quarter t the values for the 3rd month of the previous quarter $t-1$ are available for the following eight variables: ISM, SUPDEL, ORDERS, EMPLOY, HOURS, SP500, TBILL

Table 1: Monthly indicators

Name	Description (transformation)	Availability[1]
ISM	the ISM index (overall) for manufacturing (level)	1st week
EMPLOY	payroll employment (log-change)	1st week
SUPDEL	the ISM index for supplier delivery times (level)	1st week
ORDERS	the ISM index for orders (level)	1st week
HOURS	avg weekly hours of production workers (log-change)	1st week
SP500	S&P500 index (log-change)	1st week
TBILL	the 3-month Treasury bill rate (level)	1st week
TBOND	the 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	industrual production (log-change)	2nd week
STARTS	housing starts (log-level)	2nd week

[1] Reflects the availability of an observation for the previous month, either in the first or second week of the current month;

and TBOND. At the same time for the remaining four variables (CLAIMS, RSALES, IP, HS) only the values for the second month of that quarter $t-1$ have been released so far. At the next forecast origin, FO2, data availability is extended by one month. This means that for the first group of the eight variables observations for the first month of the quarter t are available and for the second group of the four variables observations for the third month of the quarter $t-1$ have been released. For each of the remaining forecast origins the data availability consequently increases by one month for all monthly variables. The dataset is organised so as to reflect the historical availability of the monthly variables in question for every month since 1985M01 until 2011M10 at the respective forecast origins.

The GDP data for a quarter $t-1$ are released in the following quarter t. The initial release takes place at the end of the first month of the quarter t, the second and third releases—at the end of second and third months of the quarter t, respectively. Such release schedule implies that at the forecast origin FO1 in a quarter t GDP data are available until the quarter $t-2$. For the forecast origins FO2 and FO3 that are within the targeted quarter t the first and second releases of GDP data for the previous quarter $t-1$ are available. Lastly, at the forecast origin FO4 that takes place in the quarter $t+1$ the third release of GDP data is known. In line with Carriero et al. (2015) we evaluate the forecasting accuracy of the econometric models using the second release of GDP data.

The actual values of the GDP growth rate published as the second release for each quarter in the forecast sample are shown in 1. Shaded areas correspond to the three recessions periods according to the NBER business cycle chronology. Note that since the NBER publishes recession dates at the monthly frequency and our forecasting exercise uses quarterly data we converted monthly NBER recession periods as follows: 1990M07—1991M03 → 1990Q3—1991Q1 (three quarters), 2001M03—2001M11 → 2001Q1—2001Q4 (four quarters) and the period of the Great Recession 2007M12—2009M06 → 2007Q4—2009Q2 (seven quarters). Thus in the forecast sample which is 107-quarter long there are 14 quarters identified as recessive periods and the remaining 93 quarters correspondingly belong to the expansion business cycle phases. The time series of the monthly indicators from the last available data vintage (2011M10) along with the recessions periods are displayed in Figure 2.

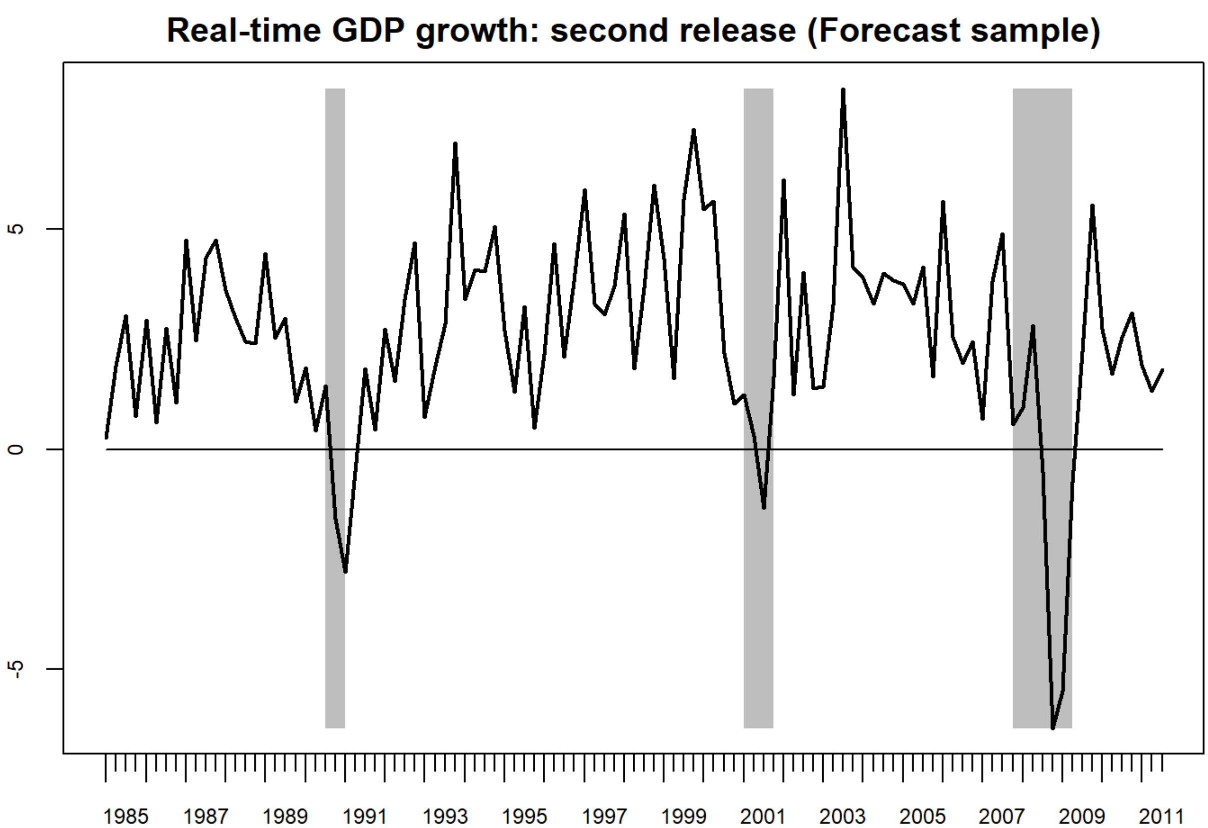


Figure 1: Data: quarterly GDP growth: second release

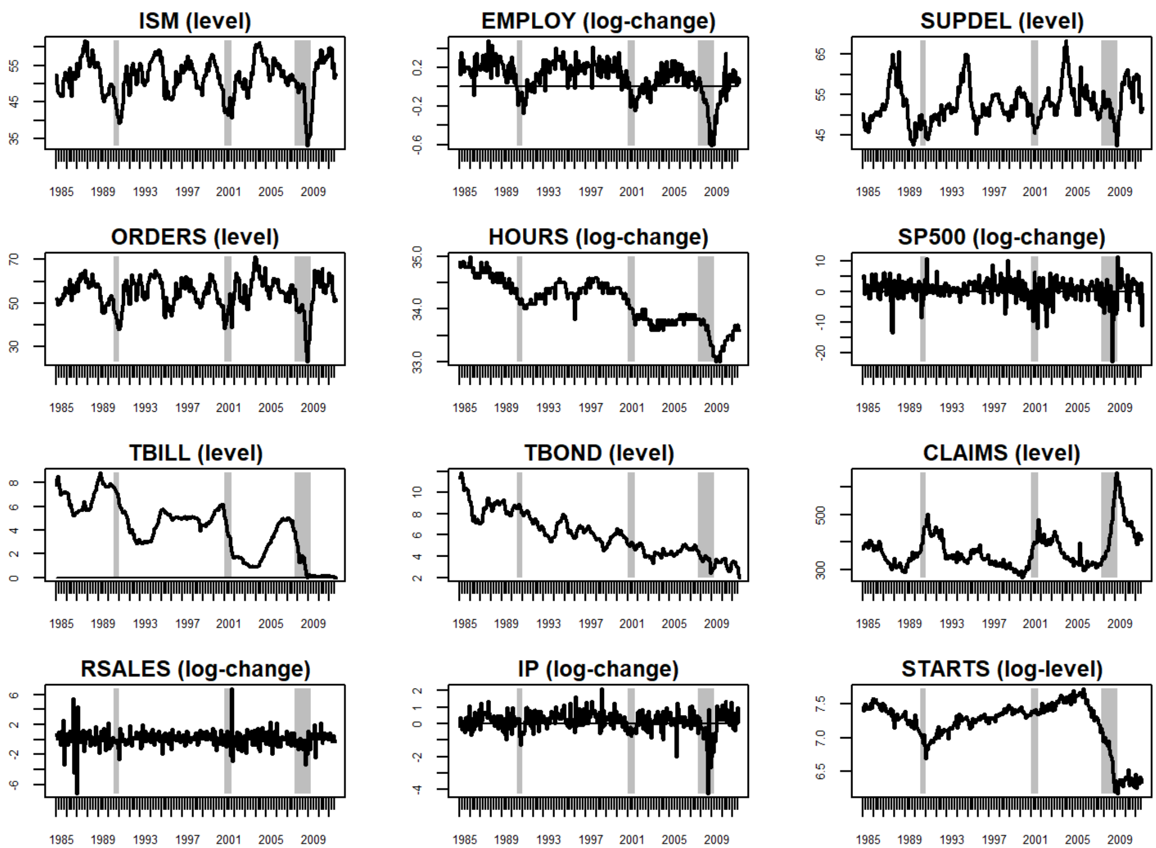


Figure 2: Data: monthly indicators, vintage from 2011-10; forecast sample

3 Econometric model of Carriero et al. (2015)

In the first subsection we describe an overall modelling framework of Carriero et al. (2015). In the second subsection we describe types of models considered in this paper.

3.1 General setup

Carriero et al. (2015) suggest the following forecasting model:

$$\begin{aligned} y_t &= X_{m,t}^{'}\beta_m + \varepsilon_{m,t} & & \\ \varepsilon_{m,t} &= \kappa_{m,t}^{0.5}\epsilon_{m,t}, & \epsilon_{m,t} &\sim NIID(0,1) \\ \ln\kappa_{m,t} &= \ln\kappa_{m,t-1} + \nu_{m,t}, & \nu_{m,t} &\sim NIID(0,\phi_m). \end{aligned}$$

The conditional mean is modelled as a linear function of the explanatory variables collected in the vector $X_{m,t}$. The sub-index $m = \{FO1, FO2, FO3, FO4\}$ corresponding to one of the forecast origins indicates that the 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 intercept, own lags of the dependent variable as well as quarterly values of the original monthly indicators such that both variables on the left- and right-hand sides of the regression equation are at the quarterly frequency. The conversion of the original monthly indicators to the quarterly frequency is achieved by skip-sampling their monthly values. For example, a monthly variable w_t is converted to the quarterly frequency by sampling each third observation of w_t in such a way that all the observations pertaining to the first, second and thirds months in every quarter are respectively collected in three quarterly time series $w_t^{(1)}, w_t^{(2)}, w_t^{(3)}$, where superscripts (i) with $i = 1, 2, 3$ indicate first, second or third months in a quarter t. In the most general form of the model, the conditional variance of the error term $\varepsilon_{m,t}$ is modelled as a time-varying stochastic process, which can also be switched off resulting in a model with constant-volatility of disturbances.

Carriero et al. (2015) use a Bayesian approach for estimation of model parameters as well as generation of out-of-sample forecasts. In doing so, the applied Bayesian shrinkage effectively deals with increased number of model parameters due to application of the skip-sampling procedure to monthly indicators. In addition, the Bayesian methodology allows generation of density forecasts that takes into account parameter estimation uncertainty as well as (potentially) time-varying variance of the error term.

Carriero et al. (2015) use normal Minnesota-style priors on the coefficient vector β_m characterised by mean zero and a diagonal covariance matrix. The degree of shrinkage is controlled by the three hyperparameters: λ_1, that determines the overall rate of shrinkage; λ_2 that sets the shrinkage rate of the monthly variables relative to that of the lags of GDP variable; and λ_3, that regulates the shrinkage rate imposed on the longer lags of regressors. The diagonal entries of the prior covariance matrix for β_m is 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 very common in the literature as acknowledged in Carriero et al. (2015, p. 845).

Carriero et al. (2015) set diffuse priors on the variance of the error term for models with constant volatility. For the models with stochastic volatility the priors on the volatility components are set independently of those set for the coeffcient 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($\ln \widehat{\kappa}_0$, 4), where $\widehat{\kappa}_0$ is the regression standard error of an AR(4) model fitted to GDP growth using available sample.

3.2 Model types

As dicussed in Carriero et al. (2015), conversion of monthly data into their quarterly counterparts by means of skip-sampling combined with the differences in release timing of the monthly indicators as well as the release schedule of GDP data results in the following information sets at each forecasting origin:

FO1: $\Omega_{m,t} = (1, y_{t-2}, y_{t-3}, W_{1,t-1}^{(1)'}, W_{1,t-1}^{(2)'}, W_{1,t-1}^{(3)'}, W_{2,t-1}^{(1)'}, W_{2,t-1}^{(2)'})'$

FO2: $\Omega_{m,t} = (1, y_{t-1}, y_{t-2}, W_{1,t}^{(1)'})'$

FO3: $\Omega_{m,t} = (1, y_{t-1}, y_{t-2}, W_{1,t}^{(1)'}, W_{1,t}^{(2)'}, W_{2,t}^{(1)'})'$

FO4: $\Omega_{m,t} = (1, y_{t-1}, y_{t-2}, W_{1,t}^{(1)'}, W_{1,t}^{(2)'}, W_{1,t}^{(3)'}, W_{2,t}^{(1)'}, W_{2,t}^{(2)'})'$,

where $W_{1,t}^{(i)}$ refers to the vector of eight variables that are released in the first week of each month and $W_{2,t}^{(i)}$ is the vector of four variables released during the second week of each month, see Table 1. Observe that for the forecast origins FO2—FO4, the lagged values of $W_{1,t}^{(i)}$ or $W_{2,t}^{(i)}$ were not included in the infomation sets.[2]

[2]The analysis with the lagged values of $W_{1,t}^{(i)}$ or $W_{2,t}^{(i)}$ is reported in the working paper version of Carriero et al. (2015), see Carriero et al. (2013).

In the following we will distinguish between the following types of models depending on what kind of regressors $X_{m,t}$ are contained in $\Omega_{m,t}$ with $X_{m,t} \subseteq \Omega_{m,t}$. First, we consider two types of the benchmark models: a historical mean model and a univariate autoregressive model of order two. We label these models as UNI-AR(0) and UNI-AR(2). Observe that the latter model was also a benchmark model in Carriero et al. (2015). For the UNI-AR(0) model the vector $X_{m,t} = (1)$ consists only of the intercept at all forecast origins and for the UNI-AR(2) model the regressor vectors are $X_{m,t} = (1, y_{t-2}, y_{t-3})'$ for FO1 and $X_{m,t} = (1, y_{t-1}, y_{t-2})'$ for FO2—FO4.

Second, we will estimate single-indicator models (SIM) by allowing one indicator at a time out of twelve available to be selected as a regressor. In sequel, we refer to such models as SIM-XXX, where XXX stands for an abbrivation used for each of the monthly indicators, see Table 1. For such models the vector of regressors at each forecast horison 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 a variable 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, we don't estimate any model that includes any of these variables at FO2.

Third, we use the single-indicator models 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) $\lambda_{m,t,i} = \sum_{s=\underline{T}}^{t-h_m} \delta^{t-h_m-s}(\widehat{e}_{m,s,i})^2$. h_m is the delay in quarters that is the specific to 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 $h_m = 1$, respectively. The combination weights are computed using the following expression:

$$\mathscr{W}_{m,t,i} = \frac{\lambda_{m,t,i}^{-1}}{\sum_{j=1}^{n} \lambda_{m,t,j}^{-1}}. \tag{1}$$

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:

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

where $\mathscr{F}_{m,t,k}$ is the nowcast density from a model $k = 1, ..., n$ for a 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 based on the logarithmic probability scores of the (past) density forecasts. These log scores are defined as the logarithm of the value of the nowcast density at the outturn, $\ln \mathscr{F}_{m,t,j}(y_t)$. The recursive weights are determined as follows:

$$\mathscr{W}_{m,t,k} = \frac{\exp\left[\sum_{s=\underline{\tau}}^{t-h_m} \ln \mathscr{F}_{m,s,k}(y_s)\right]}{\sum_{j=1}^{n} \exp\left[\sum_{s=\underline{\tau}}^{t-h_m} \ln \mathscr{F}_{m,s,j}(y_s)\right]}, \tag{2}$$

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.

Finally, we replicate the estimation results of Carriero et al. (2015) using two versions of their model: the small model that is based on five monthly regressors (ISM, EMPLOY, IP, RSALES, HS) and the larger version based on all twelve monthly regressors. We refer to these models based on the five and twelve monthly regressors as CCM-SML and CCM-LRG, respectively. The regressor vectors for these models at each forecast origin have similar structure:

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

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

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

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

but it is worthwhile noting that for the smaller model CCM-SML the vectors $W_{1,t}^{(i)}$ and $W_{2,t}^{(i)}$ contain only (ISM, EMPLOY) and (IP, RSALES, HS) variables, respectively. As mentioned above, the size of the coefficient vector β_m varies with the forecast horizon. For example, for the CCM-LRG model it takes the dimensions of 34, 10, 22 and 34 for each forecast origin FO1—FO4, correspondingly.

4 Evaluation metrics

The evaluation metrics that we use for assessment of forecast accuracy of individual models are the Root Mean Squared Forecast Error (RMSFE) for point forecasts

$$RMSFE = \sqrt{\frac{\sum_{t=\underline{\tau}}^{\overline{\tau}} (y_t - \widehat{y}_t)^2}{T}} \tag{3}$$

and the Average Logarithmic Score (ALS) for density forecasts

$$ALS = \frac{\sum_{t=\underline{\tau}}^{\overline{\tau}} \ln \mathscr{F}_t(y_t)}{T}, \tag{4}$$

where T is a number of observations in the forecast evaluation sample $[\underline{\tau}, \overline{\tau}]$ and an actual outturn is y_t and a model forecast is $\widehat{y}_t$. Since these evaluation metrics does not depend on a forecast origin we omitted the corresponding sub-index, m.

For the pairwise comparison of one model's forecast accuracy relative to that of another model we use the Relative RMSFE (RRMSFE) and the Average Logarithmic Score Difference (ALSD) for point and density forecasts, respectively. The RRMSFE and ALSD are defined as follows for each pair of models 1 and 2

$$RRMSFE_{2/1} = \frac{RMSFE_2 - RMSFE_1}{RMSFE_1} \tag{5}$$

$$ALSD_{2/1} = ALS_2 - ALS_1. \tag{6}$$

Positive/negative values of $\text{RRMSFE}_{2/1}$ indicate that on average model's 2 forecasting performance is worse/better than that of model 1 as larger RMSFE indicate less precise point forecasts. Numerical values of $\text{RRMSFE}_{2/1}$ show relative improvement/deterioration in RMSFE of model 2 relative to that of model 1. Similarly, positive/negative values of $\text{ALSD}_{2/1}$ state that on average model's 2 forecasting performance is better/worse than that of model 1, since larger values of the average logarithmic scores indicate more accurate density forecasts.

In addition to evaluation metrics based on the averaged forecasting performance over the forecast evaluation sample, we use the following recursive measures of forecast accuracy: the Cumulated Sum of Squared Forecast Error Difference (CSSFED) and the Cumulated Sum of Logarithmic Score

Difference (CSLSD). Denoting $e_{1,t}$ and $e_{2,t}$ forecast errors of models 1 and 2 in period t, respectively, the CSSFED is computed as follows

$$CSSFED_{[\underline{\tau},\overline{\tau}],1/2} = \sum_{s=\underline{\tau}}^{t}(e_{1,s}^2 - e_{2,s}^2) \quad for \quad t \in [\underline{\tau}, \overline{\tau}]\,, \tag{7}$$

where $[\underline{\tau}, \overline{\tau}]$ denotes the time interval chosen for comparison of models' forecast accuracy. The CSLSD is defined as follows

$$CSLSD_{[\underline{\tau},\overline{\tau}],2/1} = \sum_{s=\underline{\tau}}^{t}(\ln \mathscr{F}_{s,2}(y_s) - \ln \mathscr{F}_{s,1}(y_s)) \quad for \quad t \in [\underline{\tau}, \overline{\tau}]\,. \tag{8}$$

The recursive measures of forecast accuracy dissect models' forecasting performance observation by observation illustrating how the relative forecasting performance evolves over time. As a result, both CSSFED and CSLSD deliver a time sequence of cumulative differentials as opposite to aggregated measures of the forecasting performance that delivers a single 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/CSLSD indicates that forecasting gains of the model 2 over the model 1 slowly accrue over time, i.e., for the squared errors of model 1 tend to be marginally but systematically *larger* than those of model 2 and, correspondingly, the log-scores of model 1 tend to be marginally but systematically *smaller* than those of model 2. By the same token, a CSSFED/CSLSD sequence that evolves horizontally would indicate 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 observed for the most of observations. From the practical forecasting point it is of a special interest to detect such periods when one model suddenly displays worsening in its forecasting accuracy relative to its competitor. Namely, because these large discrepancies in models' forecasting performance often turn out to be very influential in computation of aggregate forecast accuracy evaluation metrics, e.g. RMSFE or ALS, commonly used for model ranking.

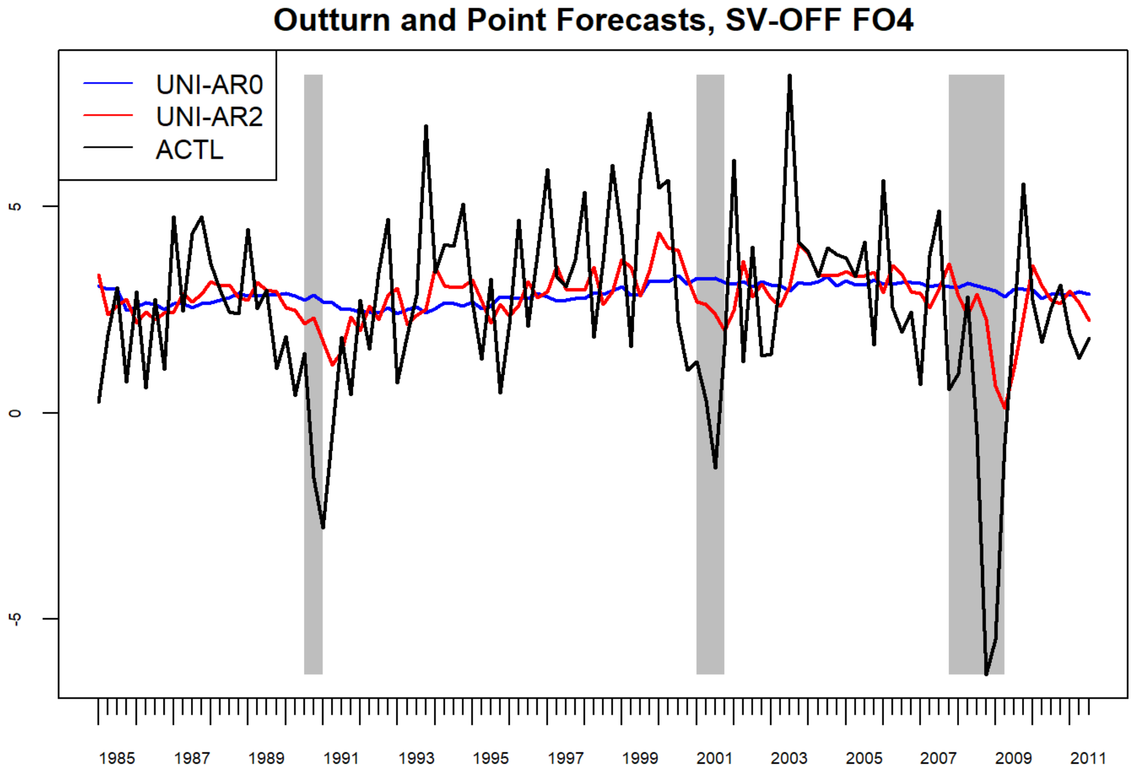


Figure 3: Point forecasts from benchmark models, w/o SV

5 Point forecasts

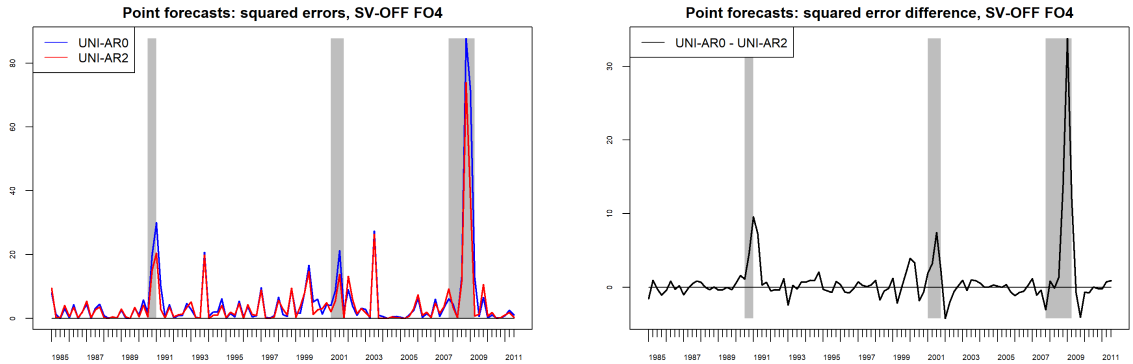
5.1 Choice of a benchmark model

When comparing models in terms of point forecast accuracy Carriero et al. (2015) used an univariate AR(2) model with constant volatility as a benchmark model. In this subsection, 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 AR(0) or the historical mean model.

The actual values of GDP growth together with forecasts from the UNI-AR2 and UNI-AR0 models produced are shown in Figure 3.[3] The forecasts of the UNI-AR2 model fluctuate around those of the UNI-AR0 model. Notably, it is easy to recognise that in times of recessions the UNI-AR2 model produces more accurate forecasts as its forecast tend to be below of the UNI-AR0 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.

Further information on the comparative forecasting performance of these two models is provided

[3] In order to keep the discussion consise 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.



(a) Squared forecast errors

(b) Squared forecast error difference: UNI-AR0 vs UNI-AR2

Figure 4: Squared point forecast errors from UNI-AR2 and UNI-AR0 models and their difference

in the left and right panels of Figure 4. In the left panel squared forecast errors for both models are shown making it is evident that in times of recessions their magnitude tends to be larger than in times of expansions. In the lower 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 RMSFE for the whole forecasting sample are mainly driven by those observations pertainig 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.

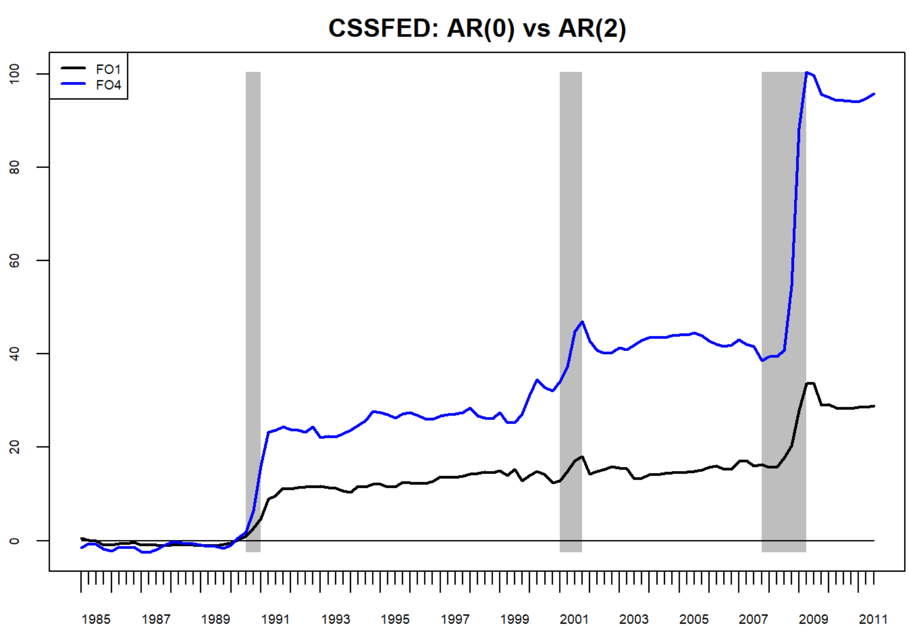


Figure 5: Point forecast accuracy, w/o SV: CCSFED$_{UNI-AR0/UNI-AR2}$

An additional information on relative forecasting performance in terms of point forecasts is provided by the CSSFED displayed in Figure 5 for FO1 and FO4. In between-recession periods the CSSFED mostly displays flat horizontal movements. As discussed in Section 4 this dynamics conforms with the idea that during expansions there are no systematic difference between forecast accuracy of these two models. In contrast, during recessions there are noticable jumps in the CSSFED indicating that the UNI-AR(0) model produces much larger squared forecast errors than the UNI-AR(2).

Table 2: Point forecast accuracy, univariate benchmark models: w/o SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$\text{RRMSFE}_{UNI-AR0/UNI-AR2}$			
	Full sample							
UNI-AR2	2.237	2.102	2.081	2.074	0.000	0.000	0.000	0.000
UNI-AR0	2.296	2.289	2.259	2.280	0.027	0.089	0.086	0.099
	Boom sample							
UNI-AR2	1.710	1.705	1.692	1.691	0.000	0.000	0.000	0.000
UNI-AR0	1.713	1.716	1.695	1.712	0.002	0.006	0.002	0.013
	Bust sample							
UNI-AR2	4.339	3.802	3.751	3.726	0.000	0.000	0.000	0.000
UNI-AR0	4.562	4.526	4.461	4.499	0.051	0.190	0.189	0.207

The results of the formal comparison in terms of RMSFE and RRMSFE are reported in Table 2 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.

Second, during expansions values of RMSFE of the UNI-AR0 model are comparable to those of the UNI-AR2 model. It is only during recessions when the difference between models' RMSFE values becomes non-negligible with the UNI-AR(2) 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 UNI-AR2 model with respect to that of the UNI-AR0 model during expansions and, conversely, to understate it during recessions. For example, from the Relative RMSFEs reported in the right panel of Table 2 one can deduce that the UNI-AR0 model produces RMSFE that is higher than that of the UNI-AR2 model by up to 10% when these RMSFE are computed over the whole forecast evaluation sample. At the same time, during booms the UNI-AR0 model produces RMSFE that is higher than that of the

UNI-AR2 model at most by about 1%. During busts the UNI-AR0 model produces RMSFE that is higher than that of the UNI-AR2 model up to about 20%.

Lastly, the fact that the UNI-AR2 model excels over the UNI-AR0 model only during recessions and that both bencmark model produce very similar forecast accuracy during expansions implies that when comparing forecasting performance of the quarterly US GDP growth rate against the benchmark UNI-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*).

5.2 Single-indicator models

The forecasting performance of the single-indicator models (SIMs) against the benchmark UNI-AR(2) model is summarised in Table 3.[4] Here again we report RMSFE and RRMSFE 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 single-indicator models there is a strong evidence of asymmetric forecasting performance during booms and busts. In absolute value the RMSFEs are higher during busts than booms. We also notice that the forecasting performance vary 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 3. When evaluated for the whole sample the three SIMs (ISM, EMPLOY, ORDERS) display the largest gains in the forecasting accuracy relative to the UNI-AR(2) among all twelve 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 UNI-AR(2) model is about 5-6% during expansions versus approximately 29-36% during recessions.

It is worthwhile mentioning that during recessions the RMSFE of the housing starts variable (STARTS) is comparable to those 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.

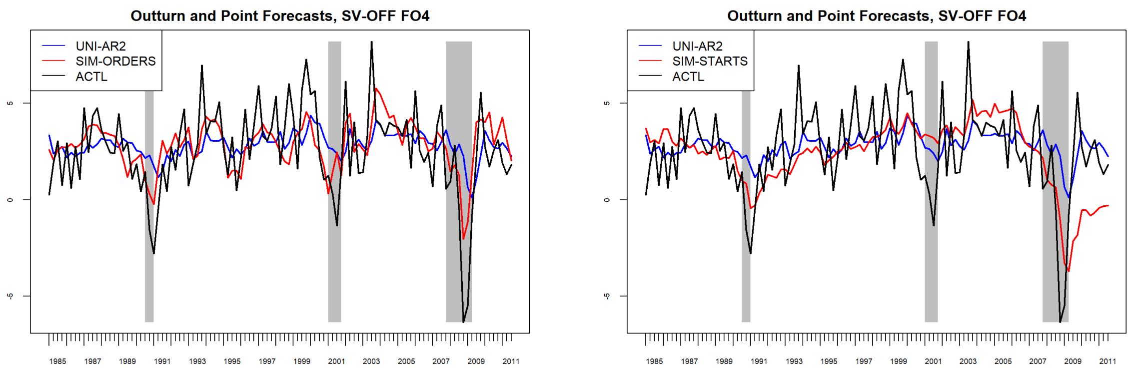
[4] The choice of the AR(2) model as a benchmark model is intentional to make our results directly comparable to those reported in Carriero et al. (2015).

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 3: Point forecast accuracy, single-indicator models: w/o SV

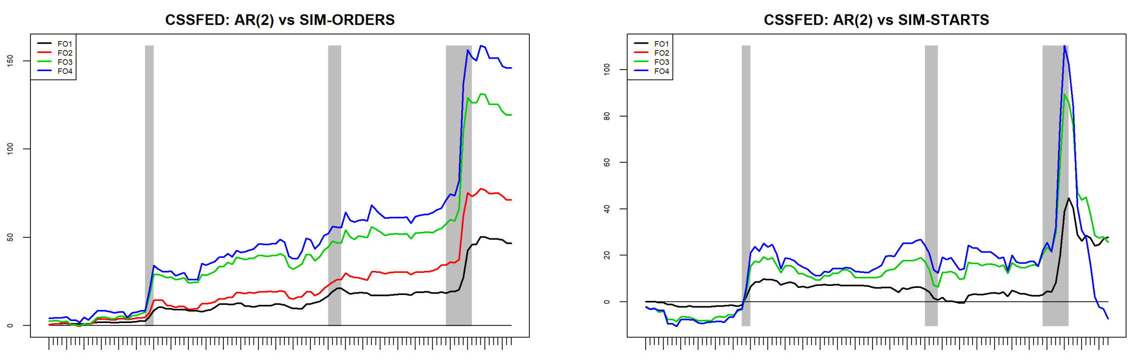
	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$\text{RRMSFE}_{SIM-XXX/UNI-AR2}$			
					Full sample			
UNI-AR2	2.237	2.102	2.081	2.074	0.000	0.000	0.000	0.000
SIM-ISM	2.172	1.960	1.870	1.775	-0.029	-0.068	-0.101	-0.144
SIM-EMPLOY	2.198	2.012	1.913	1.835	-0.017	-0.043	-0.081	-0.115
SIM-SUPDEL	2.234	2.094	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			
UNI-AR2	1.710	1.705	1.692	1.691	0.000	0.000	0.000	0.000
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			
UNI-AR2	4.339	3.802	3.751	3.726	0.000	0.000	0.000	0.000
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

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 UNI-AR(2) model in Figure 6 for FO4 and Figure 7 for FO1, FO3-FO4, respectively. These two figures show that indeed most of the forecasting gains relative to the benchmark model accumulate during recessions.



(a) SIM-ORDERS vs UNI-AR2 (b) SIM-STARTS VS UNI-AR2

Figure 6: Actual outturns and point forecasts for selected single-indicators models



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

Figure 7: CSSFED computed for selected single-indicators models

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 2002-2011 period. In the pre-crisis period the CSSFED displayed in Figure 7(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 6(b) the SIM-STARTS forecasts in this period severily underestimate the GDP growth rate at FO4.

In order to gauge how forecasting accuracy of the SIMs measured by the RMSFEs changes

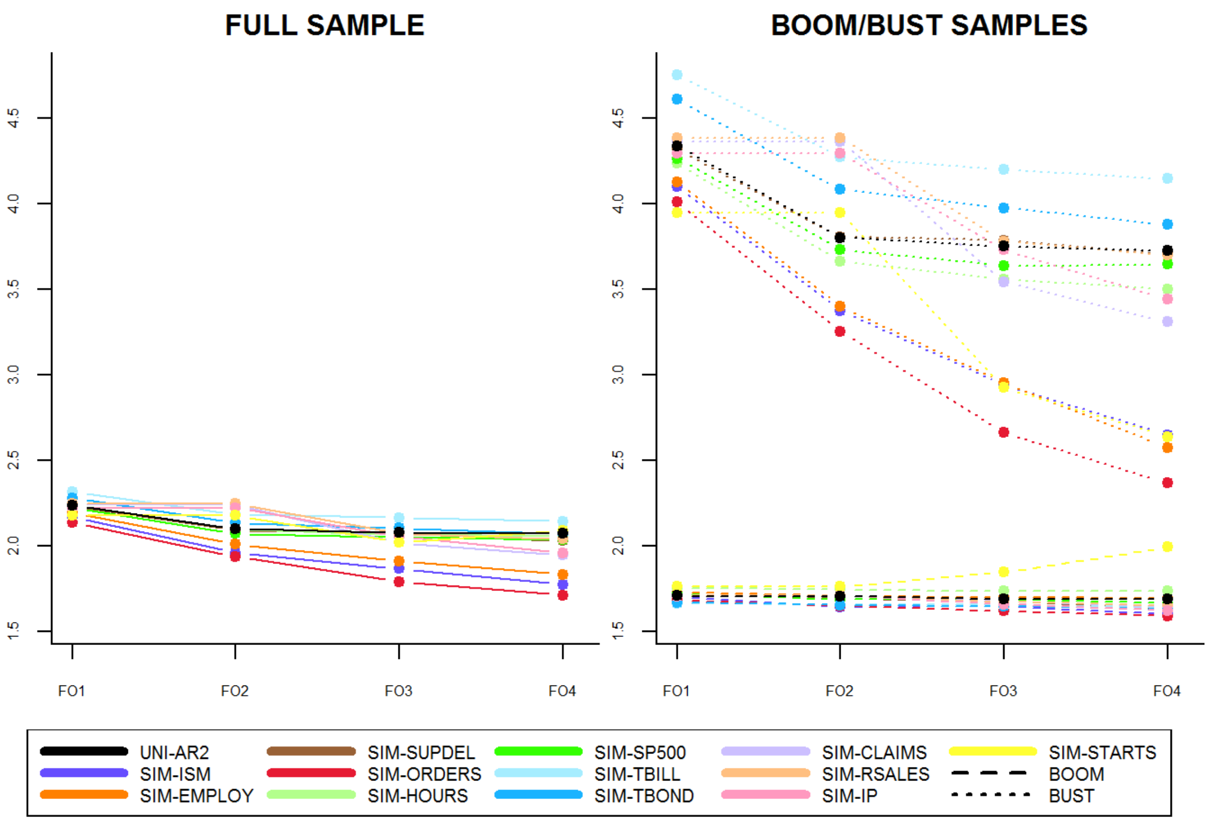


Figure 8: RMSFE summary for single-indicator models, w/o SV

with the forecast origin we present Figure 8. In the left panel the evolution of the RMSFE is shown for the whole sample as it was done, for example, in Giannone et al. (2008). For the most of the SIMs the picture is quite standard to the forecasting literature: forecast accuracy tends to increase as more information flows into the forecasting models. This standard-to-the-literature way of depicting evolution of the forecast accuracy to be contrasted with a similar plot in the right panel of the figure. In this panel it is shown how the SIM-specific RMSFE changes with the forecast origin when it is evaluated during booms (dashed lines) and busts (dotted lines). The following comments can be made based on the figure. First, as discussed above, the gap in forecasting accuracy between booms/busts clearly can be seen in this figure. Second, there is also a noticable difference on how forecasting accuracy changes with the forecast origin. During expansions the gains in forecasting accuracy displayed by the SIMs both in absolute terms and in terms relative to the benchmark model are barely noticable. It is during the busts when additional observations bring about information resulting in more accurate forecasts. This is more pronounced for these three indicators (ISM, EMPLOY, ORDERS) and for the housing starts (STARTS) than for the remainder of the indicators.

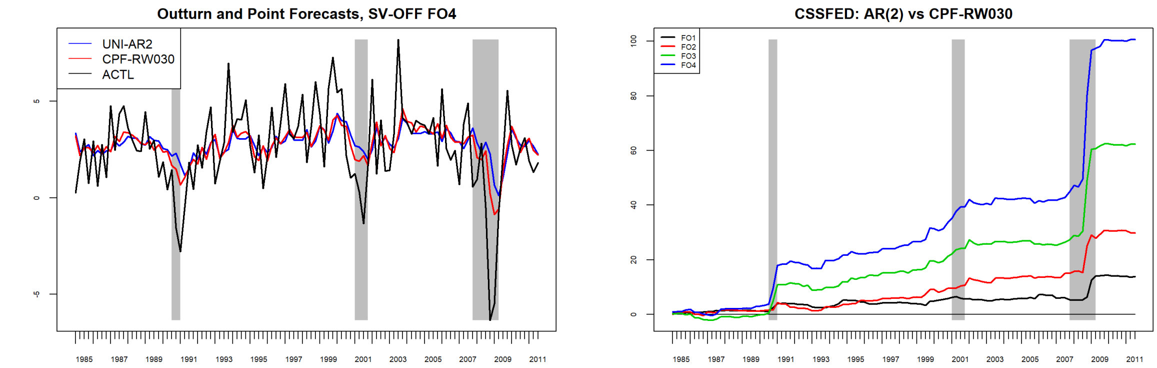
Table 4: Point forecast accuracy, point forecast combination of SIMs: w/o SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$\text{RRMSFE}_{CPF-XXX/UNI-AR2}$			
				Full sample				
UNI-AR2	2.237	2.102	2.081	2.074	0.000	0.000	0.000	0.000
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				
UNI-AR2	1.710	1.705	1.692	1.691	0.000	0.000	0.000	0.000
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				
UNI-AR2	4.339	3.802	3.751	3.726	0.000	0.000	0.000	0.000
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

5.3 Combinations of single-indicator models

The results of the forecasting exercise using combinations of single-indicator models based on their point forecasting performance is summarised in Table 4 in terms of RMSFE and RRMSFE. In general, the conclusions drawn from Table 3 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 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 single-indicator models, see Table 3. However, this is not the case during recessions: the forecasting performance of model combinations is somewhat lower than that of the best SIMs. Out of model combinations the most accurate forecasts during recessions were produced by combinations based on recursive weighting with a rather heavy discounting, $\delta = 0.30$, see Section 3.2. It is worthwhile noticing that the accuracy of forecasts based on the equal weighting were found to be inferior to that based on the recursive weighting when it is evaluated during recessions.

The actual and forecast values for the best forecasting model combination approach are shown in Figure 9(a) at FO4. The corresponding CSSFEDs for all forecast origins FO1-FO4 are shown in



(a) Outturns and forecasts, FO4 (b) CCSFED$_{AR(2)/CPF-RW030}$

Figure 9: Assessment of point forecast accuracy: UNI-AR2 and CPF-RW030

Figure 9(b). The shape of CSSFED for model combinations is very similar to that for the model SIM-ORDERS in Figure 7(b), i.e. jumps occur precisely during recessions.

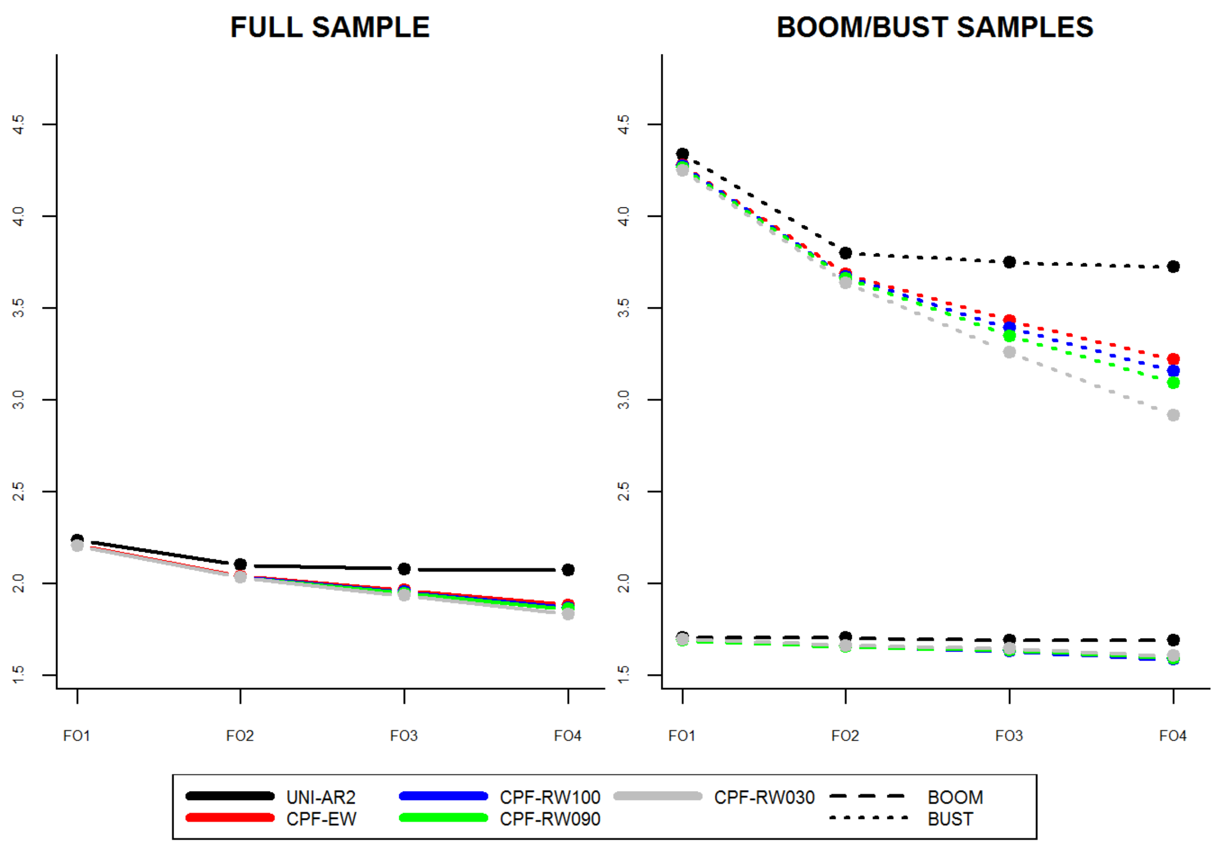
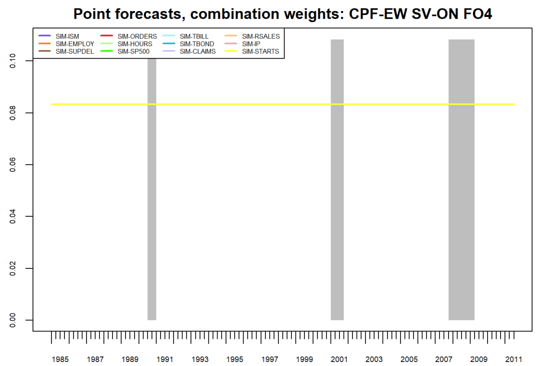
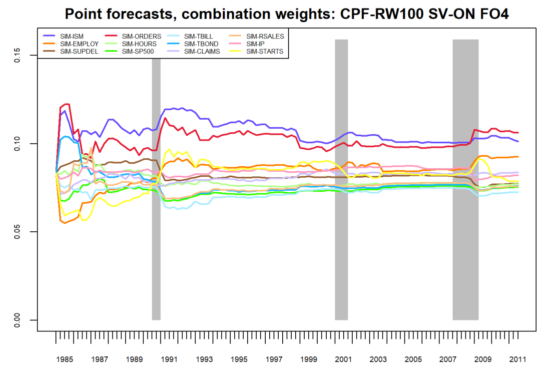


Figure 10: RMSFE summary for SIM combinations, w/o SV

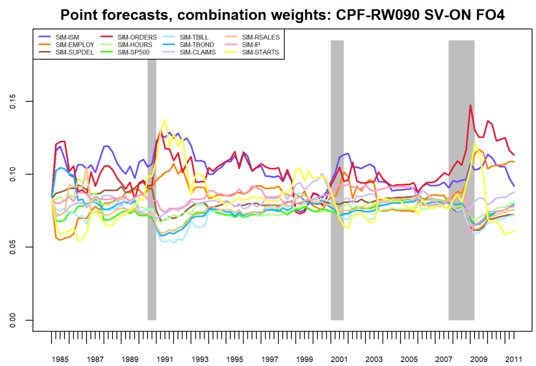
In the left and right panels of Figure 10 we show the evolution of the RMSFEs with the forecast origin when these are evaluation for the whole sample and when the evaluation is done separately for expansions(dashed lines) and recessions (dotted lines), respectively. The following comments



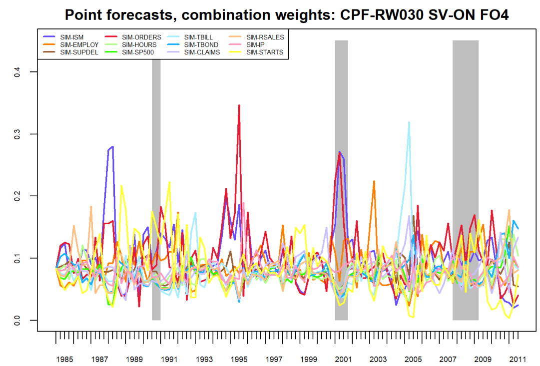
(a) CPF-EW, FO4



(b) CPF-RW100, FO4



(c) CPF-RW090



(d) CPF-RW030

Figure 11: Point forecast combinations: weights at FO4, w/o SV

can be made based on the figure. First, the model combinations display very little reduction in the reported RMSFEs with changing forecast origin during booms. Second, consistent with the results reported for the individual models most of the evidence bringing down average squared forecast errors accrues during busts. 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.

Weights of the model combinations for the latest forecast origin FO4 are shown in Figure 11. Comparing Figure 11(a) and 11(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 discount parameter from $\delta = 1$ to $\delta = 0.9$ and further to $\delta = 0.3$ increases volatility in the weight magnitude though none of the models get assigned a weight approaching value of 100%.

5.4 Models of Carriero et al. (2015)

The results for multiple-indicator models proposed in Carriero et al. (2015) based on their point forecasting performance are summarised in Table 5 in terms of RMSFE and RRMSFE. The forecasting accuracy reported for CCM-SML and CCM-LRG with constant volatility 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 above, the differences in forecasting accuracy relative to the benchmark model also can be observed for multiple-indicator models of Carriero et al. (2015). In expansions, the improvement of in RMSFE is mostly pronounced at FO4 reaching about 7% for CCM-SML and about 4% for CCM-LRG. At the earlier forecast origins both models barely improve upon the forecasting accuracy of the benchmark model. However, it is during recessions when pooling of the indicators into a single model brings about substantial gains in the forecasting accuracy. It is also interesting to observe that this improvement is much larger than that produced either by any SIM individually or by their model combinations. For example, the CCM-LRG model at FO4 yields reduction in RMSFE of about 58% compared either to 36% generated by the best SIM (SIM-ORDERS), see Table 3, or to 22% by the best model combination (CPF-RW030), see Table 4.

The finding that the gap in the RRMSFEs reported for booms and busts is larger than that observed for individuals SIMs and their combinations implies that the biases introduced by reporting RMSFE for the full forecast period for the CCM-SML and CCM-LRG models are also magnified. As a result their performance is exaggerated to a larger extent during expansions and consequently more understated during recessions.

Table 5: Point forecast accuracy, models of Carriero et al. (2015): w/o SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	RMSFE				$\text{RRMSFE}_{CCM-XXX/UNI-AR2}$			
					Full sample			
UNI-AR2	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			
UNI-AR2	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			
UNI-AR2	4.339	3.802	3.751	3.726				
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 dynamics of RMSFEs for the CCM-SML and CCM-LRG models across forecast origins is shown in Figure 12. In the left panel the evolution of RMSFEs is shown for the full sample and in the right panel its dissection according to business cycle phases. While the left panel reports a standard results of decreasing RMSFEs as the forecast origin moves forward in time, the most interesting information is contained in the right panel. Hence it is instructive to compare the right panels of Figures 12 and 10 that show the RMSFE dynamics for the multiple-indicator models of Carriero et al. (2015) and for combinations of the SIM models, respectively. Since that models of Carriero et al. (2015) as well as the SIM combinations utilise the same information set at each forecast origin it is interesting to gauge upon whether data or model pooling produces most accurate forecasts. First, we observe qualitatively similar situation when with every next forecast origin the expanding information set contributes mostly during busts to amelioration in models' forecast accuracy. Second, there is an important difference between Figures 12 and 10. While the gap between the RMSFEs in expansions and recessions still exists for the single-indicator model combinations, for the multiple-indicator models this gap eventually closes as early as at FO3 for the CCM-LRG model and it practically closes for CCM-SML at FO4. This means that at these forecast origins the forecasting accuracy of the CCM models is about of the same magnitude both in booms and busts. This result contrasts that reported in Chauvet and Potter (2013, Abstract) that "... there is a large differene forecast performance across business cycle phases. In particular, it is much harder to forecast output growth during recessions than during expansions". As a consequence, as soon as the information set is large enough to equalise forecasting performance across business cycle phases there is no need to consider different models for expansions and recessions as advocated in Chauvet and Potter (2013).

In our case, the asymmmetry remains at the earlier forecast origins: at FO1—FO3 for CCM-SML and at FO1—FO2 for CCM-LRG. For these two models it is only difference in the relative forecasting accuracy with respect to the benchmark model that remain at all forecast origins as shown in the right panel of Table 5. As argued above this difference is necessary to acknowledge when reporting results of a forecasting exercise.

The actual outturns and forecast values for CCM-SML and CCM-LRG at FO4 as well as the corresponding CSSFEDs for all forecast origins are shown in Figure 13. Observe that the CCM-LRG model that uses all twelve monthly indicators is able to accurately predict the depth of recession in the early 1990s and of the Great Recession.

5.5 Influence of stochastic volatility on point forecast accuracy

In this section we will briefly discuss the influence of introducing stochastic volatility into model on the out-of-sample forecasting accuracy. For this purpose we replicated the forecasting exercise using all the models which point forecasting performance was discussed in Sections 5.1—5.4 but this time adding the stochastic volatility. The difference in the RMSFE between the models without stochastic volatility and those with stochastic volatility are presented in Table 6. In general we find no systematic evidence that adding stochastic volatility improves accuracy of point forecasts. The magnitude of the effect in most cases comprises a couple of percentage points in either direction depending on model type and forecast origin. Similarly, the maximum effect is about 6-7 percentage points difference in the RMSFE which can be also positive as well as negative. This finding is consistent with the results reported in Table 2 of Carriero et al. (2015, p. 849) that also compare forecasting accuracy of models with constant and stochastic volatility.

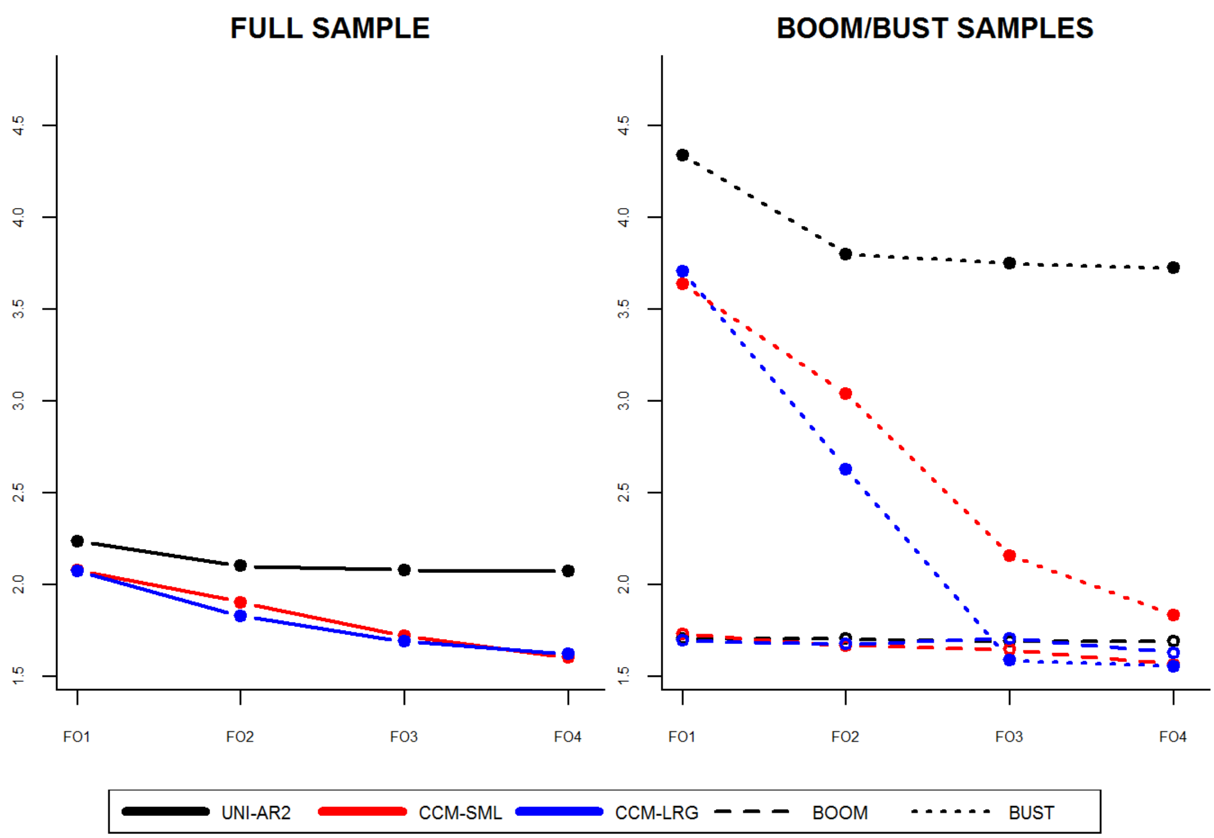
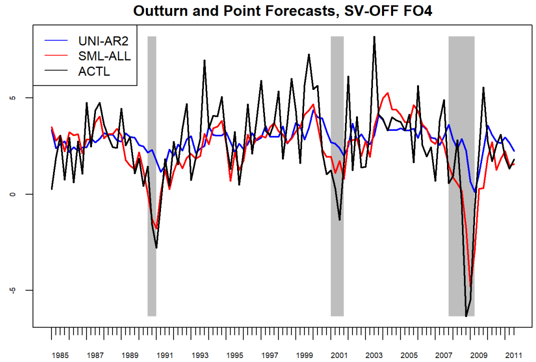
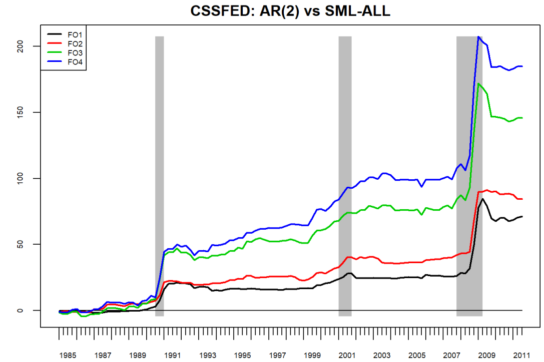


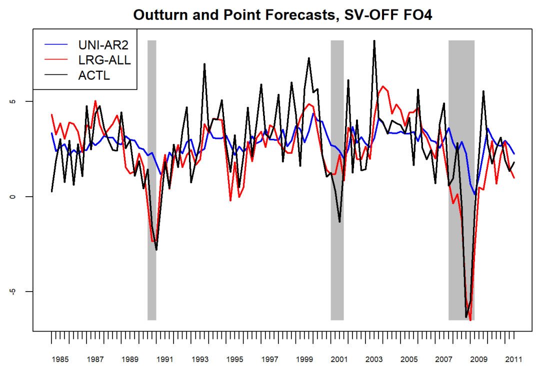
Figure 12: RMSFE summary for CCM models, w/o SV



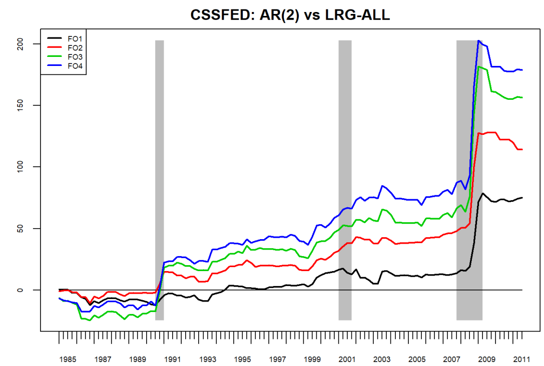
(a) CCM-SML



(b) $\text{CCSFED}_{AR(2)/CCM-SML}$



(c) CCM-LRG



(d) $\text{CCSFED}_{AR(2)/CCM-LRG}$

Figure 13: Assessment of point forecast accuracy: CCM models

Table 6: Influence 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**			
UNI-AR2	0.010	0.002	0.005	0.003	0.015	0.011	0.015	0.008	0.004	-0.010	-0.009	-0.004
UNI-AR0	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
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

Table entries are $\text{RRMSFE}_{XXX-SV/XXX}$ where the abbreviation XXX stands for a model in the corresponding row names, i.e. UNI-AR0 or CPF-EW, etc.

6 Density forecasts

6.1 Choice of a benchmark model

When comparing models in terms of density forecast accuracy Carriero et al. (2015) used an univariate AR(2) model with stochastic volatility as a benchmark model. In this subsection, we conduct an additional analysis for the choice of the benchmark model by comparing the AR(2) and AR(0) models in terms of density forecast accuracy. Average values of log scores and their difference for the models in question are presented in Table 7. 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 AR(0) model. When comparing the relative forecasting performance of these two models, then it is noticable that the $\text{ALSD}_{UNI-AR0-SV/UNI-AR2-SV}$ values are much more negative in recessions than in expansions. This indicates that the lion share of the evidence in favour of the AR(2) model over the AR(0) model stems from recession periods.

Table 7: Density forecast accuracy, univariate benchmark models: w/ SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$\text{ALSD}_{UNI-AR0-SV/UNI-AR2-SV}$			
	Full sample							
UNI-AR2	-2.253	-2.166	-2.159	-2.182				
UNI-AR0	-2.294	-2.261	-2.237	-2.239	-0.041	-0.096	-0.078	-0.057
	Boom sample							
UNI-AR2	-2.092	-2.043	-2.044	-2.059				
UNI-AR0	-2.106	-2.091	-2.090	-2.088	-0.014	-0.048	-0.046	-0.029
	Bust sample							
UNI-AR2	-3.324	-2.982	-2.926	-3.004				
UNI-AR0	-3.543	-3.393	-3.218	-3.245	-0.219	-0.412	-0.292	-0.241

An additional information on relative forecasting performance of univariate models in terms of density forecasts is provided by the CSLSD displayed in Figure 14. 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 a practically horizontal movements, there are periods characterised by upwards trends, especially shortly before or shortly after the recessions, and there are some periods with (slightly) downwards trending dynamics. Such difference in the CSLSD dynamics during booms and bust explains differences in the ALS values reported in Table 7 supporting the conclusion that most of the evidence driving differences in the forecasting accuracy of the competing models is due to observations during recessions.

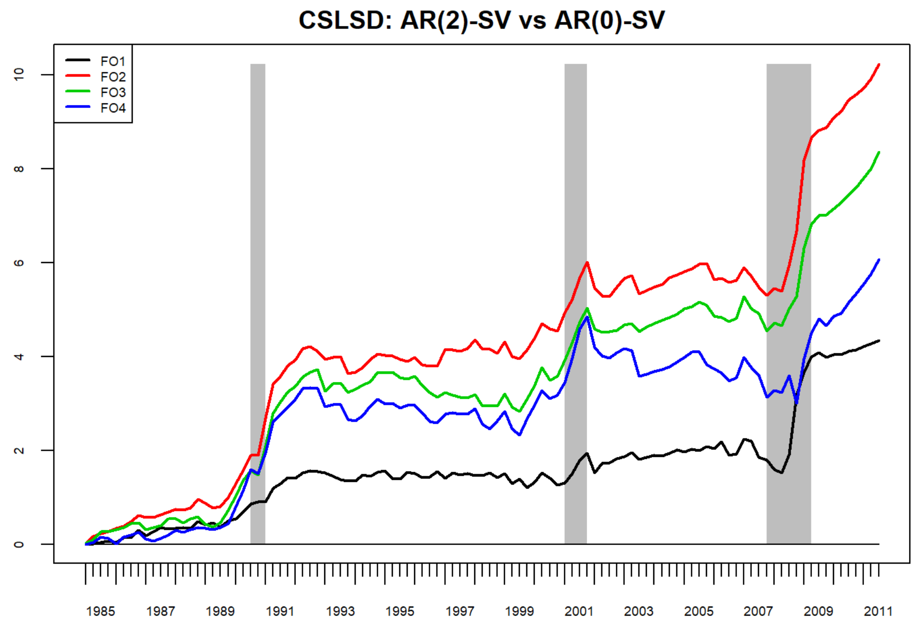
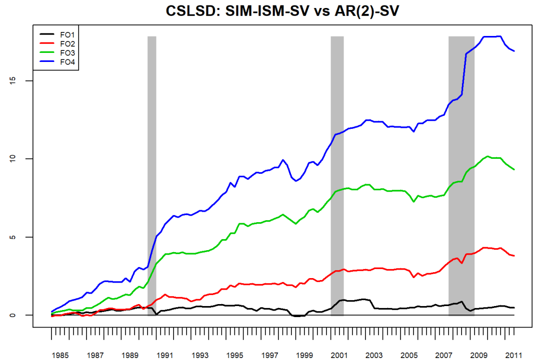


Figure 14: $\text{CSLSD}_{UNI-AR(2)/UNI-AR(0)}$: Denstity forecast accuracy, w/ SV

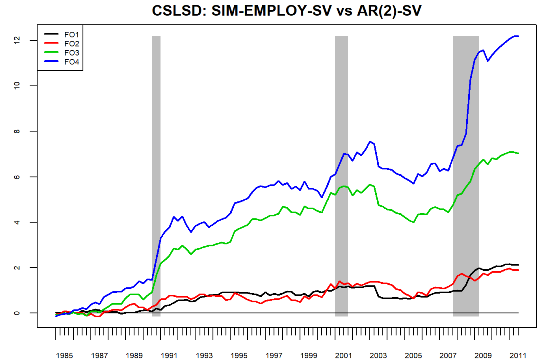
6.2 Single-indicator models

The forecasting performance of the single-indicator models is summarised in Table 8 in terms of ALSs and ALSDs against the benchmark UNI-AR2-SV model for the full sample as well as for its two sub-samples. Comparing these outcomes with the results reported in subsection 5.2 for point forecasts a number of striking similarities can be noticed. First, we also observe heterohe-neous 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 15. Fourth, the backcasts made at the forecast origin FO4 produce most accurate density forecasts compared with those made at the earlier forecast horizons.

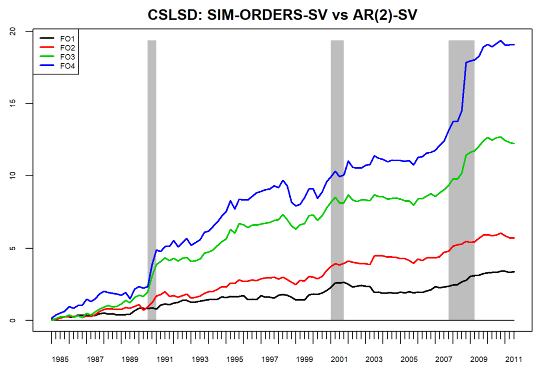
A summary over the forecasting performance of the SIMs over forecast origins is presented in Figure 16. The left panel contains the results for the full sample. Here the standard results are present: accuracy of density forecasts tends to improve as the forecast origins advance in time. In



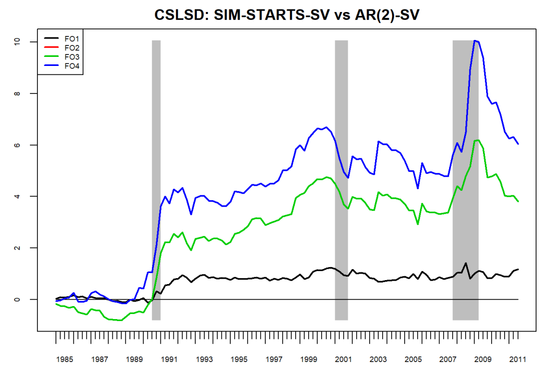
(a) $\text{CSLSD}_{SIM-ISM-SV/UNI-AR2-SV}$



(b) $\text{CSLSD}_{SIM-EMPLOY-SV/UNI-AR2-SV}$



(c) $\text{CSLSD}_{SIM-ORDERS-SV/UNI-AR2-SV}$



(d) $\text{CSLSD}_{SIM-STARTS-SV/UNI-AR2-SV}$

Figure 15: Assessment of density forecast accuracy: single-indicator models

the right panel density forecast performance is presented separately for expansions (dashed lines) and recessions (dotted lines). The advantages of SIMs over the benchmark model are much more pronounced during recessions than expansions.

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

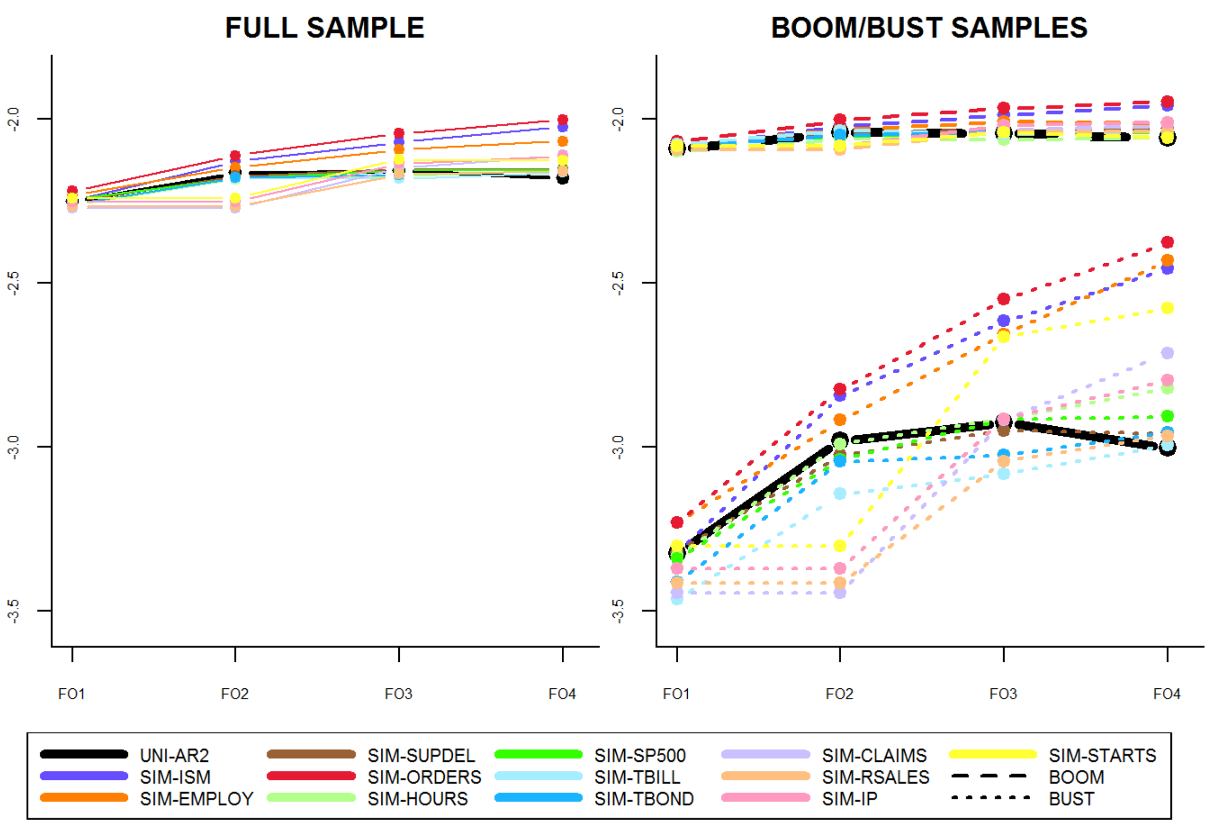


Figure 16: ALS summary for single-indicator models, w/ SV

Table 8: Density forecast accuracy, single-indicator models: w/ SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$\text{ALSD}_{SIM-XXX-SV/UNI-AR(2)-SV}$			
				Full sample				
UNI-AR2	-2.253	-2.166	-2.159	-2.182				
SIM-ISM	-2.249	-2.130	-2.072	-2.024	0.004	0.036	0.087	0.158
SIM-EMPLOY	-2.233	-2.148	-2.094	-2.068	0.020	0.018	0.066	0.114
SIM-SUPDEL	-2.251	-2.175	-2.154	-2.153	0.002	-0.010	0.006	0.029
SIM-ORDERS	-2.221	-2.112	-2.045	-2.004	0.032	0.053	0.114	0.178
SIM-HOURS	-2.262	-2.184	-2.175	-2.157	-0.009	-0.018	-0.015	0.025
SIM-SP500	-2.247	-2.182	-2.158	-2.156	0.007	-0.016	0.001	0.026
SIM-TBILL	-2.260	-2.181	-2.183	-2.164	-0.007	-0.015	-0.023	0.019
SIM-TBOND	-2.259	-2.180	-2.170	-2.159	-0.006	-0.015	-0.011	0.023
SIM-CLAIMS	-2.272		-2.150	-2.112	-0.019		0.009	0.070
SIM-RSALES	-2.268		-2.170	-2.160	-0.015		-0.010	0.023
SIM-IP	-2.252		-2.137	-2.115	0.001		0.023	0.068
SIM-STARTS	-2.242		-2.124	-2.126	0.011		0.036	0.056
				Boom sample				
UNI-AR2	-2.092	-2.043	-2.044	-2.059				
SIM-ISM	-2.087	-2.023	-1.990	-1.959	0.005	0.020	0.054	0.100
SIM-EMPLOY	-2.083	-2.032	-2.009	-2.014	0.009	0.011	0.035	0.045
SIM-SUPDEL	-2.090	-2.047	-2.034	-2.032	0.002	-0.005	0.010	0.027
SIM-ORDERS	-2.069	-2.005	-1.969	-1.948	0.022	0.038	0.075	0.111
SIM-HOURS	-2.101	-2.062	-2.063	-2.057	-0.009	-0.019	-0.019	0.002
SIM-SP500	-2.082	-2.053	-2.044	-2.043	0.010	-0.010	0.000	0.015
SIM-TBILL	-2.079	-2.036	-2.047	-2.039	0.013	0.007	-0.003	0.020
SIM-TBOND	-2.086	-2.050	-2.041	-2.039	0.006	-0.007	0.003	0.019
SIM-CLAIMS	-2.095		-2.034	-2.021	-0.003		0.010	0.037
SIM-RSALES	-2.095		-2.038	-2.038	-0.003		0.006	0.021
SIM-IP	-2.084		-2.020	-2.012	0.008		0.024	0.047
SIM-STARTS	-2.082		-2.042	-2.058	0.009		0.002	0.001
				Bust sample				
UNI-AR2	-3.324	-2.982	-2.926	-3.004				
SIM-ISM	-3.324	-2.844	-2.617	-2.458	-0.000	0.138	0.309	0.547
SIM-EMPLOY	-3.232	-2.918	-2.657	-2.432	0.092	0.063	0.269	0.572
SIM-SUPDEL	-3.320	-3.025	-2.951	-2.960	0.004	-0.043	-0.025	0.044
SIM-ORDERS	-3.231	-2.825	-2.551	-2.378	0.093	0.156	0.375	0.626
SIM-HOURS	-3.333	-2.991	-2.914	-2.821	-0.009	-0.009	0.012	0.183
SIM-SP500	-3.340	-3.039	-2.916	-2.908	-0.016	-0.058	0.010	0.096
SIM-TBILL	-3.466	-3.144	-3.083	-2.996	-0.141	-0.163	-0.157	0.008
SIM-TBOND	-3.411	-3.045	-3.026	-2.956	-0.087	-0.064	-0.100	0.048
SIM-CLAIMS	-3.446		-2.918	-2.714	-0.122		0.008	0.290
SIM-RSALES	-3.416		-3.044	-2.967	-0.091		-0.118	0.037
SIM-IP	-3.372		-2.914	-2.797	-0.048		0.012	0.207
SIM-STARTS	-3.302		-2.664	-2.577	0.022		0.262	0.427

Table 9: Density forecast accuracy, density forecast combination of SIMs: w/ SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$\text{ALSD}_{CDF-XXX-SV/UNI-AR2-SV}$			
				Full sample				
UNI-AR2	-2.253	-2.166	-2.159	-2.182				
CDF-EW	-2.247	-2.157	-2.123	-2.093	0.007	0.009	0.037	0.090
CDF-RW	-2.241	-2.133	-2.068	-2.026	0.012	0.033	0.091	0.156
				Boom sample				
UNI-AR2	-2.092	-2.043	-2.044	-2.059				
CDF-EW	-2.081	-2.033	-2.018	-2.005	0.010	0.010	0.026	0.054
CDF-RW	-2.081	-2.019	-1.988	-1.966	0.011	0.023	0.056	0.092
				Bust sample				
UNI-AR2	-3.324	-2.982	-2.926	-3.004				
CDF-EW	-3.343	-2.982	-2.819	-2.677	-0.019	0.000	0.107	0.327
CDF-RW	-3.308	-2.886	-2.598	-2.423	0.016	0.096	0.328	0.581

6.3 Combinations of single-indicator models

The results of the forecasting exercise using combinations of single-indicator models based on their density forecasting performance is summarised in Table 9 in terms of ALS and ALSD. In Figure 17 the evolution of the ALS is shown across forecast origins. 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 shown in the right panel of Figure 17 as the forecast origin moves forward the increase in the forecasting accuracy is much more pronounced during recessions than expansions. 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. The CSLSDs for the SIM combinations are shown in Figure 18. 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 upwards trending CSLSDs.

The weights attached to every SIM are shown in Figure 19 for all forecast origins. A close examination exlains 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 (2) for combinations of density forecasts is much more aggressive than that based on the discounted MSFE in Equation (1) for combinations of point forecasts, see Figure 11 for a comparison.

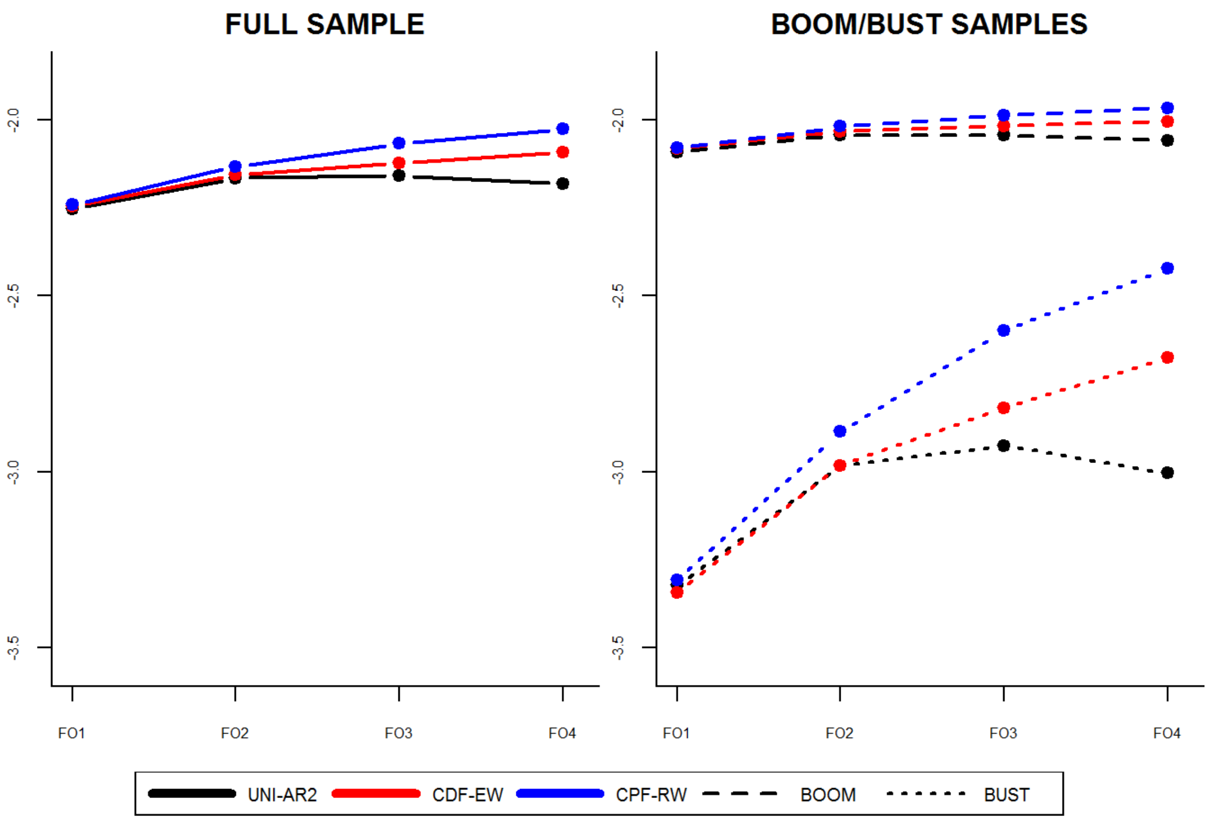
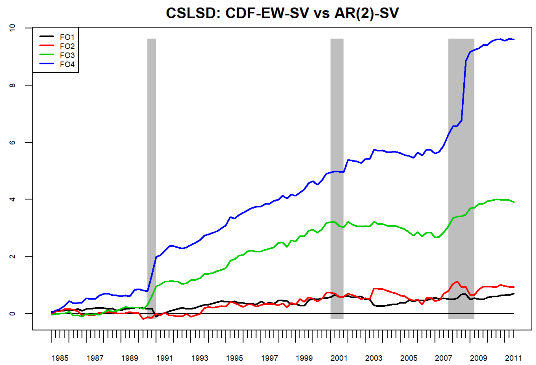
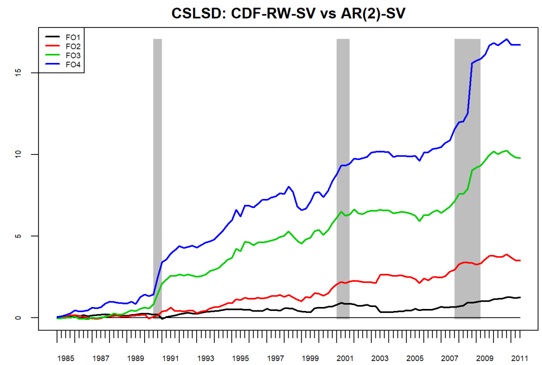


Figure 17: ALS summary for combinations of single-indicator models, w/ SV

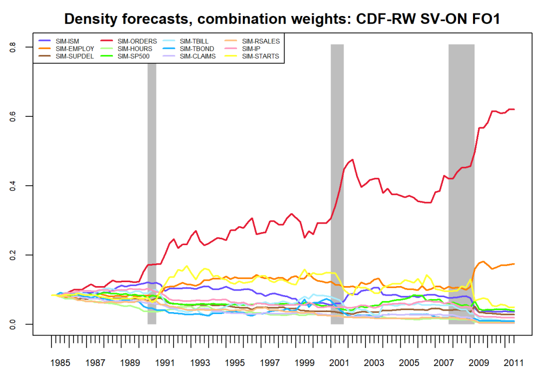


(a) CSLSD$_{CDF-EW-SV/UNI-AR2-SV}$

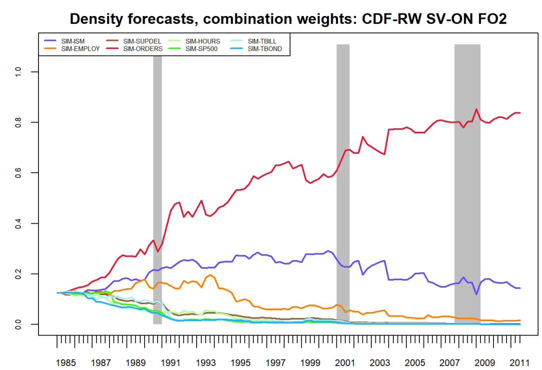


(b) CSLSD$_{CDF-RW-SV/UNI-AR2-SV}$

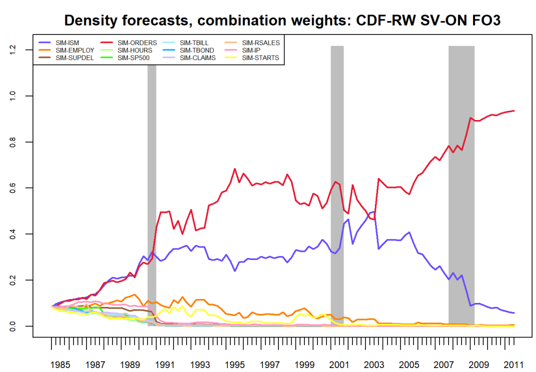
Figure 18: Assessment of density forecast accuracy: density forecast combinations of SIMs



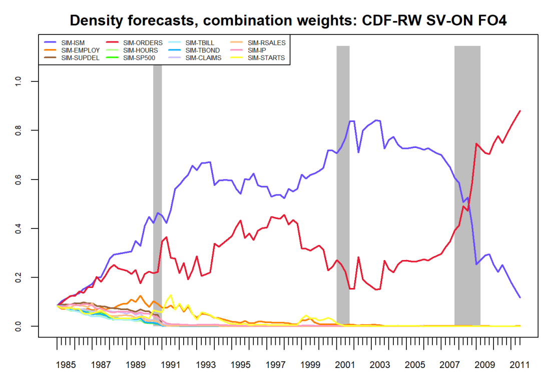
(a) FO1



(b) FO2



(c) FO3



(d) FO4

Figure 19: Density forecast combinations: recursive weights, w/ SV

Table 10: Density forecast accuracy, models of Carriero et al. (2015): w/ SV

	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	ALS				$\text{ALSD}_{CCM-XXX-SV/UNI-AR2-SV}$			
				Full sample				
UNI-AR2	-2.253	-2.166	-2.159	-2.182				
CCM-SML	-2.207	-2.090	-1.965	-1.891	0.046	0.076	0.195	0.292
CCM-LRG	-2.165	-2.063	-1.959	-1.913	0.088	0.103	0.200	0.270
				Boom sample				
UNI-AR2	-2.092	-2.043	-2.044	-2.059				
CCM-SML	-2.054	-2.001	-1.926	-1.873	0.038	0.042	0.118	0.185
CCM-LRG	-2.043	-1.996	-1.956	-1.915	0.049	0.047	0.088	0.144
				Bust sample				
UNI-AR2	-3.324	-2.982	-2.926	-3.004				
CCM-SML	-3.223	-2.682	-2.223	-2.006	0.101	0.300	0.703	0.998
CCM-LRG	-2.977	-2.508	-1.976	-1.898	0.347	0.474	0.950	1.107

6.4 Models of Carriero et al. (2015)

The results for multiple-indicator models proposed in Carriero et al. (2015) are summarised in Table 10. For the full sample these results are very close to those reported for the corresponding models *Small BMFSV* and *Large BMFSV* in Table 3 of Carriero et al. (2015, p. 852). A summary of the forecasting performance of these models in terms of ALS across forecast origins is shown in Figure 20: The left panel reports the results for the full sample and the results for recessions and expansions are reported in the right panel. The results reported for the full sample coincide with those reported in Carriero et al. (2015), i.e. more information increases accuracy of density forecasts. However, it is instructive to compare the right panels of Figures 20 and 12 as well as Tables 10 and 5 as there is a number of similarities in the performance of these models in terms of point and density forecasts. First, at the earlier forecast origins ALS are lower during busts than booms. However, the gap between the reported ALSs across business cycle phases practically closes at FO4 for CCM-SML and as early as at FO3 for CCM-LRG. This means that at these forecast origins absolute numerical accuracy of density forecasts is similar in expansions and recessions. Second, the convergence that takes place in models' forecast accuracy across business czcle phases is mainly due to improvement in the forecast accuracy during recessions. There is some improvement in ALS that takes place also during expansions but this improvement is much less pronounced. Third, the asymmetry in the forecasting performance during booms and busts relative to the benchmark model still remains and it tendentially grows with the forecast origin as more information becomes available. Fourth, pooling information in one model results in better density forecasts than pooling single-indicator models: A conclusion that is very similar to that

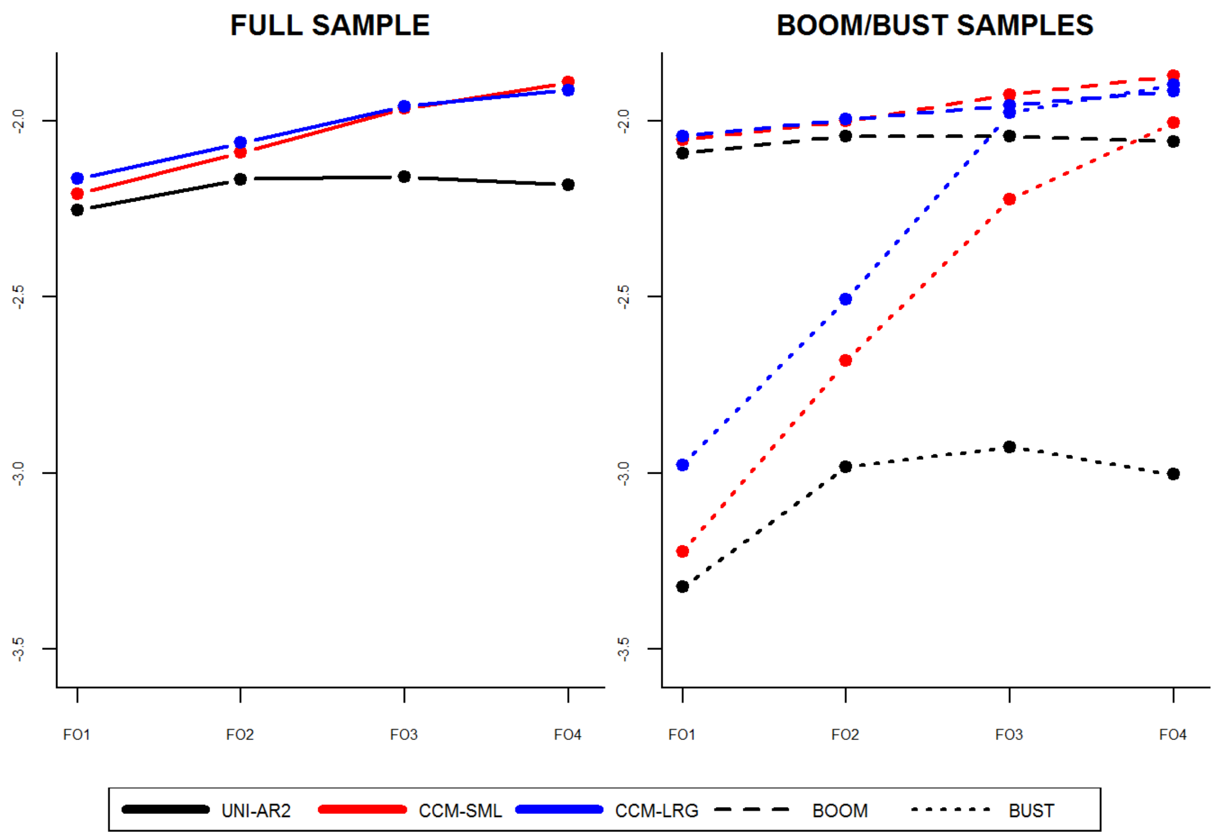


Figure 20: ALS summary for combinations of Carriero et al. (2015), w/ SV

for point forecasts. Fifth, the asymmetry observed in the relative density forecasting peformance implies that by reporting ALSD for the full period results in a biased evaluation of the models' forecasting accuracy: it is exaggerated during expansions and consequently understated during recessions relative to the forecasting performance of the benchmark model.

6.5 Influence of stochastic volatility

The conclusion that there is no systematic effect on accuracy of point forecasts by adding stochastic volatility to the forecasting models was stated in Section 5.5. 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.

The effect of the stochastic volatility on the density forecast accuracy is reported in Table 11. Entries in the table are differences in ALS reported between models with stochastic and constant volatility in the residual error term. Positive entries imply that on average density forecast accuracy of the models with stochastic volatility was higher, negative entries indicate the opposite. As seen, for the full sample as well as for expansions the models with stochastic volatility produce higher

Table 11: Influence of stochastic volatility on density forecast accuracy

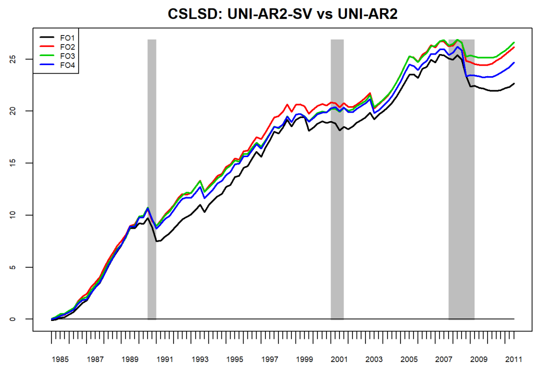
	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4	FO1	FO2	FO3	FO4
	Full sample				**Boom sample**				**Bust sample**			
UNI-AR2	0.212	0.245	0.249	0.231	0.297	0.311	0.309	0.300	-0.356	-0.199	-0.154	-0.227
UNI-AR0	0.182	0.212	0.228	0.231	0.293	0.300	0.299	0.299	-0.556	-0.375	-0.240	-0.225
SIM-ISM	0.193	0.231	0.238	0.244	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
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

ALS than their conterparts 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, SIMs and SIM combinations there are negative entries at all forecast origins FO1—FO4 indicating that during recessions models with constant volatility produced on average more accurate density forecasts. For the two remaining models (CCM-SML and CCM-LRG) there is negative difference in ALS at FO1—FO2. At the later forecast origins, FO3—FO4, this difference is positive.

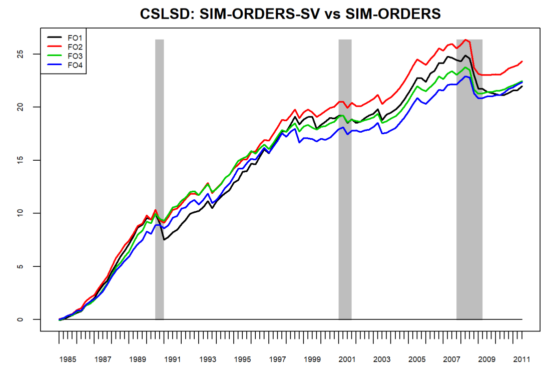
In tracking the reason for this at the first glance surprising result it is instructive to inspect plots of CSLSD for the models with stochastic and constant volatility. These are displayed for the benchmark UNI-AR2 and SIM-ORDERS models as well as for the CCM-SML and CCM-LRG models in Figure 21. In general, upwards trending behaviour of the CSLSDs indicates more or less steadily gains in density forecast accuracy of models with stochastic volatility over those 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. For the CCM-SML-SV and CCM-LRG-SV models these setbacks at FO3—FO4 are less pronounced, which explains comparatively better forecasting performance over models with constant volatility during recessions at these forecast origins.

An example of density forecasts of the SIM-ORDERS models with stochastic (black curve) and constant (red curve) volatility for one selected quarter in the Great Recession (2008Q3) is presented in Figure 22. The vertical line represents the actual outturn of the GDP growth rate in this quarter. The density forecasts of the model with constant volatility have longer tails which ensures higher values of the probability scores for this realisation of GDP growth.

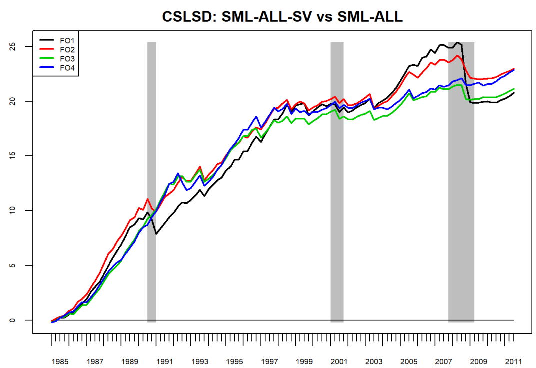
The following lesson can be drawn from this exercise when comparing density forecasts from



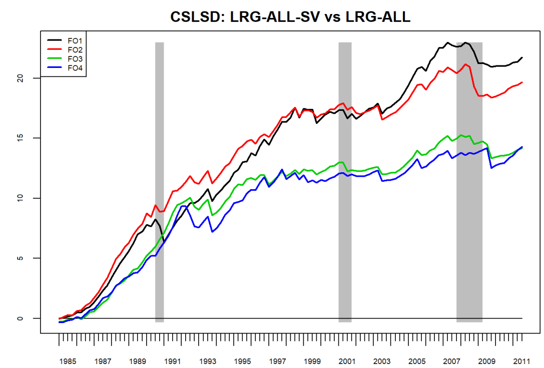
(a) $\text{CSLSD}_{UNI-AR2-SV/UNI-AR2}$



(b) $\text{CSLSD}_{SIM-ORDERS-SV/SIM-ORDERS}$



(c) $\text{CSLSD}_{CCM-SML-SV/CCM-SML}$



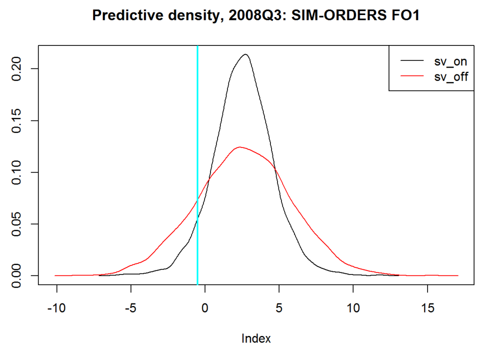
(d) $\text{CSLSD}_{CCM-LRG-SV/CCM-LRG}$

Figure 21: Effect of stochastic volatility on density forecast accuracy: CSLSD

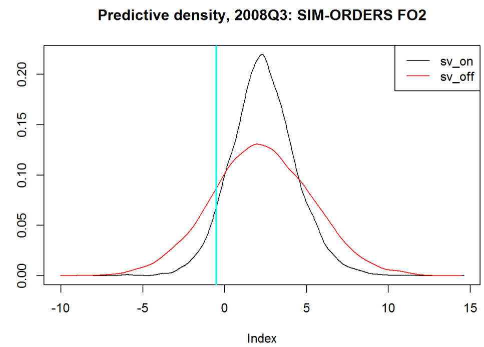
models with and without time-varying volatility of the error term. For the data at hand introducing stochastic volatility brings about more accurate density forecasts most of the times. However, there might be several outturns that lie further in tails, especially, during periods of economic downturns, when fatter tails of models with constant volatility help these models in producing relatively higher (log) predictive scores. Finally, forecast evaluation over sub-samples accompanied by measures of recursive assessment of relative forecasting performance is helpful to uncover pecularities in models' forecasting performance as well as to provide explanations for their occurence.

7 Conclusions

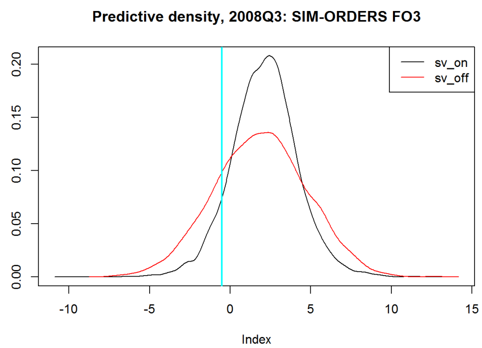
In this paper we extend the analysis of Chauvet and Potter (2013) on the differences of forecasting accuracy of macroeconometric models during recessions and expansions in several directions. First, we scrutinise the choice of a benchmark model against which the forecasting performace of more sophisticated models is usually measured. Chauvet and Potter (2013) use a univariate second-order autoregressive model as the benchmark model. The same benchmark model is also used in Carriero



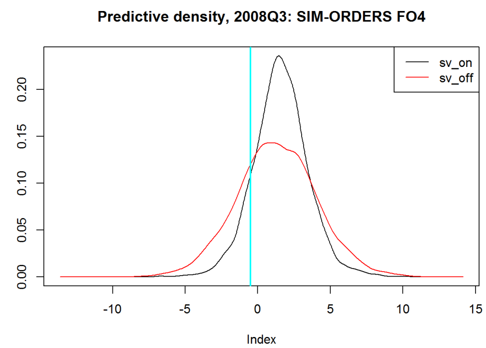
(a) FO1



(b) FO2



(c) FO3



(d) FO4

Figure 22: SIM-ORDERS density forecasts: 2008Q3

et al. (2015) which serves as a starting point for our paper by providing the data set and specification of econometric model used in our forecasting exercise. More specifically, we compare the forecasting performance of the mentioned AR(2) model with that of the model based on historical mean or AR(0) model. We find that the forecasting performance of the AR(2) model relative to that of the AR(0) model do vary with the business cycle phases. During expansions both models produce very similar forecast accuracy whereas it is only during recessions when the AR(2) produces forecasts of superior accuracy relative to those of the AR(0). Peculiar enough, this is a similar finding to that reported in Chauvet and Potter (2013) when comparing the forecasting performance of several state-of-the-art macroeconometric models relative to that of the AR(2) model. Hence during expansions forecasts based on the historical mean serve as a good benchmark for assessing the additional gains in the forecasting performance brought about by more sophisticated models.

Second, we complement the analysis of Chauvet and Potter (2013) that reviewed the quality of forecasts from macreconometric models relying on single-frequency quarterly data by considering the forecasting performance of an econometric model that utilises mixed-frequency data, i.e., in our case data sampled both at the quarterly and monthly frequencies. The advantage of using the mixed-frequency data is that it allows us to timely incorporate monthly information into forecasts and hence monitor changes in forecast accuracy not only between also within a given quarter. Thirdly, the evaluate the forecasting performance of the models in questions using not only point but also density forecasts. Finally, within the modelling framework we investigated whether pooling the data into one model or combining multiple single-indicator models produces superior forecasts. At this stage we considered model combinations using past performance in terms of point and density forecast accuracy.

We conclude that the findings of Chauvet and Potter (2013) on the importance to distinguish between models forecasting performance during recessions and expansions largely carry over also for forecasting models that utilise mixed-frequency data. As in Chauvet and Potter (2013), we generally record less precise forecasts during recessions than during expansions. However, in our case this holds either for forecasts generated using partial information (single indicator models and their combinations) or for forecasts generated at the earlier forecast origins, i.e., when there is still no complete information about a targeted quarter in our nowcasting exercise. As soon as a forecasting model is estimated using all available information this asymmetry in forecast accuracy across the business cycle phases disappears. In this sense data pooling results in superior forecast accuracy relative to model combinations.

Remarkably, while the recessions/expansions dichotomy in *absolute* forecasting performance eventually disappears (at least, for some of our models in question) the differences in terms of the *relative* forecasting performance with respect to the benchmark models manifests itself for all models and all forecast origins. As in Chauvet and Potter (2013), we find that our mixed-frequency models perform at best only marginally better than univariate benchmark models during expansions but substantially outperform those during recessions. This result has important implications when reporting 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 preferred and, typically, more sophisticated forecasting model relative to benchmark models typically results in exaggeration of the relative forecasting accuracy of the former model in expansions and, consequently, its understatement during recessions. This delivers a biased message to anyone interested in forecasts—be it general public, academics, practitioners or policy-makers.

Last but not least, our conclusions drawn above refer to point and density forecasts alike.

References

Aastveit, K. A., C. Foroni, and F. Ravazzolo (2017). Density forecasts with MIDAS models. *Journal of Applied Econometrics 32*(4), 783–801. jae.2545.

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics 32*(1), 48–68.

Aastveit, K. A., F. Ravazzolo, and H. K. V. Dijk (2017). Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics 0*(0), 1–15.

Bates, J. and C. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 451–468.

Bulligan, G., M. Marcellino, and F. Venditti (2014). Forecasting economic activity with higher frequency targeted predictors. *International Journal of Forecasting 31*(1), 188–206.

Carriero, A., T. E. Clark, and M. Marcellino (2013). Real-time nowcasting with a Bayesian mixed frequency model with stochastic volatility. Working Paper 9312, Centre for Economic policy Research, London.

Carriero, A., T. E. Clark, and M. Marcellino (2015). Realtime nowcasting with a Bayesian mixed frequency model with stochastic volatility. *Journal of the Royal Statistical Society Series A 178*(4), 837–862.

Chauvet, M. and S. Potter (2013). Forecasting output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Forecasting*, Volume 2, pp. 1–56. Amsterdam: North Holland.

Edge, R. M., M. T. Kiley, and J.-P. Laforte (2010). A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a DSGE model. *Journal of Applied Econometrics 25*(4), 720–754.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(1), 57–82.

Giannone, D., F. Monti, and L. Reichlin (2016). Exploiting the monthly data flow in structural forecasting. *Journal of Monetary Economics 84*, 201 – 215.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Kim, H. H. and N. R. Swanson (2017). Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*.

Mazzi, G. L., J. Mitchell, and G. Montana (2014). Density nowcasts and model combination: Nowcasting Euro-Area GDP growth over the 2008-09 recession. *Oxford Bulletin of Economics and Statistics 76*(2), 233–256.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review 62*, 902–917.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics 33*(3), 366–380.

Siliverstovs, B. (2017a). International stock return predictability: On the role of the United States in bad and good times. *Applied Economics Letters 24*(11), 771–773.

Siliverstovs, B. (2017b). Short-term forecasting with mixed-frequency data: A MIDASSO approach. *Applied Economics 49*(13), 1326–1343.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies 21*(4), 1455–1508.