Forecasting Euro Area Recessions in Real-Time *

Deutsche Bundesbank

Inske Pirschel†

February 5, 2016

Abstract

I present evidence that the linear mixed-frequency Bayesian VAR provides very sharp and well-calibrated monthly real-time recession probabilities for the euro area for the period from 2004 until 2013. The model outperforms not only the univariate regime-switching models for a number of hard and soft economic indicators and their optimal linear combinations, but also a real-time recession index obtained with Google Trends data. This result holds irrespective of whether the joint predictive distribution of several economic indicators or the marginal distribution of real GDP growth is evaluated to extract the real-time recession probabilities of the mixed-frequency Bayesian VAR. The inclusion of the confidence index in industry proves to be crucial for the performance of the model.

Keywords: Density nowcasting, Real-time recession forecasting, Mixed-frequency data, Bayesian VAR, Regime-switching models, Linear opinion pool, Google Trends

JEL-Codes: C53, E32, E37

---

*The views expressed in this paper represent the author's personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank.

†inske.pirschel@bundesbank.de

1 Introduction

Forecasts of macroeconomic activity are highly important for economic policymakers' decision making processes. In addition to precise point forecasts, a reliable and timely prediction of business cycle turning points can be extremely useful for the design of appropriate economic policy, since the effectiveness of monetary and fiscal policy measures can depend heavily on the current phase of the business cycle.[1] In practice, however, there are many problems associated with the real-time availability of many macroeconomic time series. These include mixed data frequencies, the irregular and sometimes varying publication lags of various macroeconomic indicators (often referred to as *ragged edges*) and data revisions. They pose huge challenges to professional forecasters (see Giannone et al., 2008, for a detailed discussion) and should therefore be taken into account when assessing the accuracy of alternative forecasting approaches.[2]

Researchers looking to separate periods of economic expansion from recessions typically turn to non-linear regime-switching models (for recent applications see Camacho et al., 2014; Chauvet and Piger, 2008; Nalewaik, 2012). As an alternative, Bayesian density forecasting approaches (overviews are provided, for example, in Geweke and Whiteman, 2006; Karlsson, 2013) can be used to compute the probability that the economy is in a specific business cycle phase at a certain point in time. This has been documented, for example, by Österholm (2012), who estimates the probability of a recession in the US in the third and fourth quarter of 2008 with a quarterly linear Bayesian vector autoregression. Dovern and Huber (2015) estimate a linear Bayesian global vector autoregression and show that the model delivers probabilistic recession forecasts that are more precise than those obtained with country-specific models. However, the analyses in both of these papers are not conducted in a real-time setting since the models used there do not account explicitly for the aforementioned features of real-time data.

By contrast, the linear mixed-frequency Bayesian vector autoregression (MFBVAR) proposed by Schorfheide and Song (2015) is well-suited to identifying business cycle turning points in real-time, since it can be estimated on mixed-frequency data with ragged edges. The model has been proven to increase the accuracy of short-term point and density forecasts for a number of variables (Schorfheide and Song, 2015), yet it is still an open question whether it can also achieve forecast gains for the real-time detection of business cycle phases.

With this paper, I fill this gap and provide evidence that the MFBVAR provides very accurate monthly real-time recession probabilities for the euro area for the period from 2004 until 2013. The risks of a recession are defined here as the probability that current-quarter GDP growth is part of a sequence of two consecutive quarters, both displaying negative GDP growth rates. They are obtained from the joint predictive distribution of the back-, now- and forecasts

[1] Lo and Piger (2005) provide supporting empirical evidence for monetary policy and Auerbach and Gorodnichenko (2012) for fiscal policy.

[2] Recently, the success of different econometric forecasting methods in providing a reliable assessment of the prevailing economic conditions in terms of GDP growth point forecasts, while at the same time coping with the outlined difficulties, has been demonstrated. These methods include bridge equation models (Baffigi et al., 2004; ECB, 2008), MIDAS-models (Kuzin et al., 2011; Schumacher, 2014) and factor models (Banbura and Rünstler, 2011; Schumacher and Breitung, 2008) as well as combinations of the aforementioned methods (Angelini et al., 2011; Marcellino and Schumacher, 2010). The relative accuracy of these methods has been studied, for example, in Foroni and Marcellino (2014).

for euro area real GDP growth in a real-time forecasting setting. I compare the accuracy of the MFBVAR real-time recession signals with those obtained with univariate regime-switching models for a number of hard and soft economic indicators as well as their optimal linear combinations. Moreover, I consider a real-time recession index based on Google Trends data that is constructed as a population-weighted mean of the Internet query shares for the word "recession" in the eleven largest euro area countries.

Related papers that focus on the real-time detection of recessions (see Hamilton, 2011, for a comprehensive overview) often rely on monthly variables such as industrial production as a proxy of overall economic activity (Anas et al., 2008; Bellgo and Ferrara, 2009; Chauvet and Piger, 2008; Schreiber, 2014). Exceptions to this are Aastveit et al. (2014) and Camacho et al. (2014), who estimate models that account for many of the outlined features of real-time data. In particular, Aastveit et al. (2014) solve the mixed-frequency data issue by applying the Bry-Boschan rule (Bry and Boschan, 1971), an algorithm that detects recessions, to a bridge equation model nowcast and compare the accuracy of the real-time recession probabilities thus obtained to those obtained with an autoregressive Markov-switching model for Norwegian GDP. Camacho et al. (2014) estimate a mixed-frequency Markov-switching dynamic factor model for the euro area which captures not only co-movements across various economic indicators through a common business cycle factor, but also shifts in the business cycle regime.

In all these studies, the real-time recession signals are compared with an official business cycle chronology such as, for example, that established by the CEPR Euro Area Business Cycle Dating Committee or that of the NBER for the US. Accordingly, in this paper, I use the CEPR euro area business cycle turning points as a benchmark to evaluate the alternative forecasting approaches. However, while most of the aforementioned papers confine their analysis to a comparison of the official business cycle turning points to those obtained with their respective econometric models, I compute formal measures that explicitly assess the calibration as well as the sharpness of the probabilistic recession forecasts obtained with the different methods. An approach is said to deliver well-calibrated probability forecasts if the empirical event probability conditional on a forecast is close to that probability forecast, i.e. that it actually rains in 70% of the times rain was announced with a probability of 70%. Sharpness, on the other hand, refers to the question of whether the probability forecasts are clear-cut, i.e. whether they are clustered around the confident values of zero and one, rather than the ambiguous value of 0.5. The ideal probabilistic forecast maximizes sharpness subject to calibration (Ranjan and Gneiting, 2010). This implies that the real-time recession signals need to be not only very timely but also clear-cut.

Beyond that, I investigate the discriminatory skill of the different approaches. That is to say I explore the extend to which the real-time recession probabilities obtained with the alternative models are useful signals when binary forecasts for the occurrence or non-occurrence of a recession have to be issued. The ad-hoc binary event classifier typically used in related papers is 0.5, and a recession is announced if the recession probability exceeds this threshold (see, for example, Chauvet and Piger, 2008; Hamilton, 1989). However, as it turns out, this threshold is not always optimal in the sense that it maximizes the number of correct recession predictions and, simultaneously, minimizes the number of false alarms. Lahiri and Wang (2013) present a survey of different measures to evaluate probabilistic recession forecasts which take this aspect into account and I apply the receiver operating characteristic and the Peirce skill score to assess the different models' discriminatory skill. Note that these evaluation approaches are closely related to the literature on the signals approach, where potential indicators for economic crises are analyzed with respect to their early warning properties (see, for example, Boysen-Hogrefe et al., 2015; Reinhart and Kaminsky, 1999).

My findings show that the MFBVAR real-time recession probabilities are very sharp and well-calibrated and that only a univariate Markov-switching model for the confidence index in industry yields probabilistic recession forecasts that perform equally well. Both models also have the highest skill to discriminate between recessions and expansions in real-time, although the optimal binary event classifier used to translate the probabilistic forecasts into binary recession signals varies for both models. By contrast, the real-time recession signals obtained from other soft indicators such as the Economic Sentiment Indicator or the confidence index in retail sales are much less well-calibrated. In fact, these methods deliver many recession signals in non-recession periods, which would suggest that they are potentially driven by more than economic fundamentals. The probabilistic forecasts obtained with the models for the hard economic indicators, in particular for industrial production and real GDP, on the other hand, lack sharpness due to the long publication lag of the respective data. As a consequence, they have no discriminatory skill to distinguish between recession and expansion periods in real-time. The combinations of the probabilistic forecasts of the univariate regime-switching models improve upon most of their components in all dimensions considered here. However, even when an optimal combination scheme is applied, the pooled real-time recession probabilities are outperformed by those of the MFBVAR. The Google Trends real-time recession indicator performs better than most univariate regime-switching models and pools, but it is clearly worse than the MFBVAR in terms of calibration, sharpness and discriminatory skill. The index delivers very ambiguous real-time recession signals particularly between the two recession periods in the sample and proves to be of very limited use.

Finally, in the robustness analysis, I provide evidence that the inclusion of the confidence index in industry is crucial for the good performance of the MFBVAR. Moreover, I investigate the extent to which the MFBVAR real-time recession signals can be improved upon by simultaneously assessing the joint development of several economic indicators through the multivariate predictive distribution of these variables, rather than just the path of GDP growth alone. My findings indicate that no significant gains in accuracy are obtained compared to the benchmark, where the real-time risks of a recession are defined as the probability that current-quarter GDP growth is part of a sequence of two consecutive quarters both displaying negative GDP growth.

The remainder of this paper is structured as follows. In section (2) I give an overview of the euro area business cycle since 2000, while in section (3) I describe the dataset used for the empirical application in this paper. In section (4) I set out the alternative forecasting approaches, which are evaluated using the formal measures described in section (5). In section (6) I present the main results, while the results of the robustness checks are shown in section (7). Finally, in section (8) I conclude.

2 The Euro Area Business Cycle

The CEPR Euro Area Business Cycle Dating Committee has been publishing business cycle turning points for the euro area since 2003.[3] Table (1) displays the euro area business cycle phases since 2000 as stated by the CEPR.

Dates	Business cycle phase
Until January 2008	Expansion
February 2008 - April 2009	Recession
May 2009 - July 2011[4]	Expansion
August 2011 - January 2013	Recession
Since February 2013	Expansion

Table 1: CEPR euro area business phases since 2000.

The committee defines a recession as "... a significant decline in the level of economic activity, spread across the economy of the euro area, usually visible in two or more consecutive quarters of negative growth in GDP, employment and other measures of aggregate economic activity for the euro area as a whole, and reflecting similar developments in most countries. A recession begins just after the economy reaches a peak of activity and ends when the economy reaches its trough." (Artis et al., 2003). In total, the committee has identified two recessions since 2000, namely the Great Recession of 2008-09 and the recession in connection with the European debt crisis of 2011-13. These are marked by the shaded areas in panel (a) of Figure (1), which displays quarter-on-quarter euro area real GDP growth since 2000. While the first recession period lasted for 15 months, the second recession in the sample persisted for 18 months in total. During the Great Recession, euro area real GDP growth turned negative in the second quarter of 2008 and remained so until the second quarter of 2009. The strongest decrease in real GDP amounted to -2.5% and occurred in the first quarter of 2009. The European debt crisis, by contrast, was much milder, with real GDP growth dipping by a maximum of -0.6% in the fourth quarter of 2012. In this recession, real GDP growth rates were negative from the fourth quarter of 2011 until the first quarter of 2013. Panel (b) of Figure (1) plots the course of euro area real GDP over as many as 10 quarters after all CEPR-dated peaks since 1970 (normalized to one). It can be seen that compared to earlier recessions in the euro area, the Great Recession was by far the most severe in terms of depth, while the recession in connection with the European debt crisis was characterized by a decline in economic activity that was comparably prolonged but only moderate overall.

[3] The publication lag for the CEPR business cycle turning points is quite substantial. For example, the euro area business cycle peak that occurred in January 2008 was not announced until 31 March 2009 only. Similarly, the trough in April 2009 was identified with a delay of more than 12 months.

[4] The CEPR has recently abandoned its practice of announcing the month of the business cycle turning point. Hence, from July 2011 onwards, I set the first month of the quarter announced as being a business cycle turning point as the month of the respective peak or trough. This assumption is quite conservative and requires the real-time recession signals of the alternative approaches to be very timely. The results of an evaluation where the second or third month of a quarter is set as the turning point are very similar to those presented in section (6) and are available upon request.

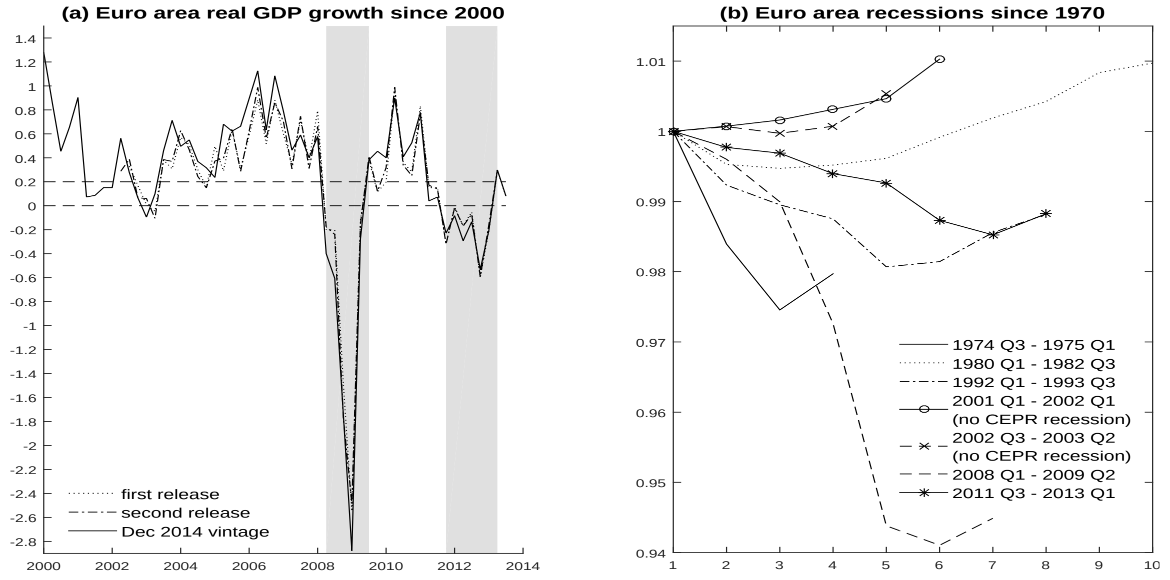


Figure 1: Euro area economic activity.

In addition to the CEPR, other authors have attempted to establish a monthly business cycle chronology for the euro area (see Anas et al., 2008; Billio et al., 2012, for recent examples that also cover the Great Recession of 2008/2009). Their assessment of the Great Recession, which Billio et al. (2012) date from September 2008 until July 2009, differs slightly from that of the CEPR committee. Moreover, there is also disagreement as to whether there was another recession in the euro area between 2000 and 2005. Billio et al. (2012) point to an industrial recession from September 2001 until May 2006, which Anas et al. (2008) date from December 2000 until November 2001. However, the view of the CEPR is that the overall evidence did not support a "fully-fledged recession but rather a prolonged pause in the growth of economic activity" (Artis et al., 2003). This is also confirmed in Panel (b) of Figure (1), which includes not only the official CEPR recessions since 1970 but also two periods between 2000 and 2005 with weak real GDP growth rates of less than 0.2%.

Note that there are other formal approaches to identifying business cycle turning points such as the well-known Bry-Broschan rule (Bry and Boschan, 1971; Harding and Pagan, 2002). For the period from 2000 onwards, however, this rule delivers the same business cycle chronology for the euro area as the CEPR Euro Area Business Cycle Dating Committee.

3 Data

For the empirical application in this paper I use a real-time dataset that consists of 123 monthly data vintages for October 2003 until December 2014, all of which start in January 1991.[5] Each of these data vintages provides a historical snapshot of the data at the beginning of each month, as

[5]The very small number of vintages that were unavailable for some variables, were replaced by the data vintage for the previous month. Moreover, since the data vintages for the unemployment rate only start after 1991, all vintages were augmented with data taken from the OECD database.

it was available at the time. This implies that the dataset reflects not only the publication lag of each variable with respect to the reference date, i.e. the date at which the snapshot was taken, but also changes in the data flow over time driven by recent improvements in the timeliness of various indicators. The dataset was obtained from the real-time database of the European Central Bank's Statistical Data Warehouse in early December 2014. A detailed description of the database, the variables included as well as the treatment of issues such as data revisions, changing variable definitions and the composition of the euro area over time can be found in Giannone et al. (2010). All series are seasonally adjusted, and natural logarithms are taken for all variables not expressed in rates.

Euro area monthly indicators	Reporting lag in January 2008
Industrial production excluding construction	3 months
CPI	2 months
Unemployment	2 months
New passenger car registrations	2 months
Money supply M1	2 months
Money supply M3	2 months
Economic sentiment indicator	1 month
Stock market index	1 month
Oil price	1 month
Confidence index industry	1 month
Confidence index retail sales	1 month

Table 2: Euro area monthly indicators with respective reporting lag in early January 2008.

The dataset includes eleven monthly indicators for the euro area which are summarized in Table (2). The reporting lag for each indicator, which is displayed in the second column of the table, illustrates the heterogeneity in the timeliness of the publication of different series. For example, in early January 2008, the most recently available observation for industrial production excluding construction was dated October 2007, while for the Economic Sentiment indicator, the figure for December 2007 was already available.

For quarterly euro area real GDP, which is also included in the analysis, the first official release is usually published about 45 days after the end of the reference quarter. Hence, the respective observation is included for the first time in the data vintage of the third month of the following quarter. This implies not only that the figure for current-quarter GDP is unknown throughout the quarter, but also that the figure for previous-quarter GDP is not available in the first and second month of a given quarter. Hence, in each quarter,, an estimate of current-quarter GDP (referred to as the *nowcast*) and in some cases of previous-quarter GDP (referred to as the *backcast*) as well have to be computed. To increase the estimation sample for the univariate quarterly Markov-switching model for real GDP growth (see section 4.3), each of the data vintages for GDP is augmented with data from the 14th update of the area-wide model database (Fagan et al., 2001) covering the period from 1970 Q1 until 1990 Q4.

4 Forecasting Approaches

To assess the real-time probabilities of a recession in the euro area, I implement a number of different econometric models which are described below in sections 4.1 to 4.3. In addition, I consider a real-time recession indicator based on Internet search data that is described in section 4.4.

4.1 The Bayesian Mixed-Frequency VAR

Consider the following monthly VAR

$$X_t = C + A_1 X_{t-1} + ... + A_p X_{t-p} + \epsilon_t, \quad (1)$$

where the vector $X_t = (x^m_{1,t}, ..., x^m_{11,t}, x^q_t)'$ contains the 11 observable monthly indicators listed in Table (2) and latent monthly real GDP x^q_t. Following Bańbura et al. (2010) I include the variables in log-levels rather than growth rates so as to not lose information that might possibly be contained in the trends. p denotes the number of lags included in the estimation and is set to $p = 6$. C is a vector of constants, $A_1, ..., A_p$ are parameter matrices, and ϵ_t is a vector of independently identically distributed white noise error terms with zero mean and covariance matrix Σ.

To account for the mixed frequencies and the ragged edges of the dataset, the VAR outlined in equation (1) is rewritten in state-space form with a time-varying measurement equation (Schorfheide and Song, 2015) that reads

$$Y_t = S_t \Lambda Z_t. \quad (2)$$

The corresponding transition equation for the states $Z_t = (X'_t, ..., X'_{t-p+1})'$ is simply the companion form of the monthly VAR described in equation (1). In equation (2) the time-varying diagonal selection matrix S_t governs that the states contained in Z_t are included in the observation vector Y_t only if they are truly observable, while the matrix Λ aggregates latent monthly real GDP into its observed quarterly counterpart. In particular, following Schorfheide and Song (2015) the log of quarterly real GDP is assumed to be observable every third month only and to be equal to the average over the three unobserved monthly GDP figures in the respective quarter, i.e. $y^q_t = \frac{1}{3}\left(x^q_t + x^q_{t-1} + x^q_{t-2}\right)$. Hence, for $t = 3, 6, 9, ..., T^b$, where T^b is the last month in which a quarterly GDP figure is observable, the observation vector reads $Y_t = (y^m_{1,t}, ..., y^m_{11,t}, y^q_t)'$, where $y^m_{j,t}$ are the $j = 1, ..., 11$ monthly indicators and y^q_t denotes observed quarterly real GDP. By contrast, in the first and second month of each quarter y^q_t is dropped from Y_t. Moreover, at the current edge, e.g. for $t > T^b$, y^q_t is never included and depending on their publication lags some of the $y^m_{j,t}$ are dropped from Y_t as well.

The mixed-frequency state-space model outlined above is estimated with Bayesian techniques using data up to month $T > T^b$. This involves the estimation of the marginal posterior distributions of the unknown VAR parameters $A_1, ..., A_p, C$ and Σ as well as the estimation of the unknown state vector $Z_{1:T}$. Following Schorfheide and Song (2015), I rely on a version of the normal inverse Wishart prior that retains the main principles of the widely used Minnesota prior (Kadiyala and Karlsson, 1997; Litterman, 1986). The prior is augmented to constrain the sum of coefficients of the VAR (Sims and Zha, 1998) as well as to incorporate the belief that the variables in the VAR follow a common stochastic trend. I implement this prior using the dummy variable approach outlined in Bańbura et al. (2010).[6]

The initial values of the state vector Z_0 are sampled conditional on a presample ranging from April 1991 until December 1994. A Gibbs sampler then iteratively samples the VAR parameters $A_1, ..., A_p, C$ and Σ as well as the unknown states $Z_{1:T}$ from their respective conditional posterior distributions.

For each of the retained Gibbs draws of the VAR coefficients $A_1^i, ..., A_p^i, C^i, \Sigma^i$ and the vector of states $Z_{1:T}^i$ a shock vector ϵ_{T+h}^i is drawn from $N(0, \Sigma^i)$ and equation (1) is iterated forward to compute forecasts for the monthly observable and unobservable variables in $\hat{X}_{T+h}^i$ with $h = 1, ..., 12$. The forecasts for unobservable monthly GDP are transformed into their quarterly counterparts based on equation (2). From these I compute the implied forecasts for quarterly GDP growth $\Delta\hat{y}_{T^B+\tilde{h}}^i$, where T^B denotes the last quarter for which GDP was observable and $\tilde{h} = 1, ..., 3$. The set of $\{\Delta\hat{y}_{T^B+\tilde{h}}^i\}_{i=1}^N$ approximates the predictive distribution of the back-, now- and forecasts of quarterly euro area GDP growth that can be used to compute pointforecasts as the mean or median of the distribution and real-time recession probabilities.

Note that depending on the current information set, i.e. the month of the quarter in which the prediction is made, $\Delta\hat{y}_{T^B+1}$ could either denote a backcast (implying that we are in the first or second month of a quarter when last quarter GDP is not available yet) or a nowcast. Correspondingly, $\Delta\hat{y}_{T^B+2}$ refers to a nowcast if it is computed in the first two months of a quarter and to a 1-quarter ahead forecast in every third month of a quarter, and so on. For example, in January 2008 the most recently available observation for GDP refers to the third quarter of 2007 ($T^B = 2007Q3$) and $T^B + 1$ denotes the backcast for the fourth quarter of 2007, while $T^B + 2$ refers to the nowcast for the first quarter of 2008. By contrast, two months later, in March 2008, the figure for the fourth quarter of 2007 has been released ($T^B = 2007Q4$), and the nowcast for the first quarter of 2008 is denoted as $T^B + 1$. This has to be taken into account in the following when computing the MFBVAR real-time recession probabilities.

According to a widely used (approximate) definition, the economy is in a recession if real GDP growth is negative for at least two consecutive quarters. I will therefore define the real-time risks of a recession as the probability that the nowcast for current-quarter GDP growth (i.e. either $\Delta\hat{y}_{T^B+1}$ or $\Delta\hat{y}_{T^B+2}$, depending on the current information set) is part of a sequence of two consecutive quarters, both displaying negative GDP growth rates. This criterion implies that the GDP growth nowcast could be either the first or the second period of a two-quarter recession sequence. Hence, taking into account the data availability in month T, the real-time recession probabilities implied by the MF-BVAR can be computed as

$$\pi_T^{MFBVAR} = \begin{cases} Pr\left(\Delta\hat{y}_{T^B} < 0,\ \Delta\hat{y}_{T^B+1} < 0 \cap \Delta\hat{y}_{T^B+1} < 0,\ \Delta\hat{y}_{T^B+2} < 0 \,|Y_T\right) & \text{for} \quad T = 3, 6, ... \\ Pr\left(\Delta\hat{y}_{T^B+1} < 0,\ \Delta\hat{y}_{T^B+2} < 0 \cap \Delta\hat{y}_{T^B+2} < 0,\ \Delta\hat{y}_{T^B+3} < 0 \,|Y_T\right) & \text{otherwise.} \end{cases} \tag{3}$$

[6] A detailed outline of the prior is provided in the appendix.

From the Gibbs sampler output π_T^{MFBVAR} can be easily obtained as

$$\pi_T^{MFBVAR} = N^{-1} \sum_{i=1}^{N} \mathcal{I}\Big(\Delta \hat{y}^i_{T^B:T^B+3}\Big), \tag{4}$$

where $\mathcal{I}(.)$ denotes an indicator function that is equal to one if, and only if, the GDP growth nowcast for the current quarter is part of a consecutive sequence of two quarters both displaying negative GDP growth. Note that in section (7) I consider alternative recession definitions to extract the real-time recession probabilities from the predictive distribution of the MFBVAR to see how far the model's performance is robust to the definition in equation (3).

4.2 A Quarterly Bayesian VAR

As a benchmark, I estimate a quarterly version of the model outlined in section (4.1) for each of the monthly data vintages. This implies that all monthly observations beyond T_b, i.e. the last month for which real GDP is available, are dropped and that all monthly indicators are aggregated to a quarterly frequency. Since the quarterly BVAR does not include any latent variables, there is no need to set up a state-space system as described above. However, apart from that the estimation procedure, the prior specification and the computation of the predictive densities and real-time recession probabilities are equivalent to those of the MFBVAR.

4.3 Markov-Switching Models

4.3.1 Univariate Markov-Switching Models

For selected indicators in the dataset, I set up the following univariate model:

$$\Delta y_t = \mu_{s_t} + \psi_{s_t} \Delta y_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \sim N(0, \sigma_{s_t}), \tag{5}$$

where Δy_t denotes the first difference of the respective indicator.[7] The latent discrete variable s_t is assumed to evolve as a two-state, 1st order Markov-switching process, i.e. $s_t = \{E, R\}$, with transition probabilities

$$P(s_t = j | s_{t-1} = i) = p_{ij}, \;\; i, j = E, R. \tag{6}$$

This model implies that the dynamics of the process described in equation (5) may differ between the two regimes E and R, thus allowing for structural breaks in the time series which can be estimated. Assuming that $\mu_E > \mu_R$ and that E are expansionary business cycle phases, while R stands for recession periods, the model can be used to identify business cycle turning points and to compute recession probabilities. In particular, the probability that the economy

[7] For the sake of simplicity I use the subscript t for both, the model in monthly frequency for the monthly indicators and the model in quarterly frequency for real GDP growth.

is in a recession in period t given the observations $y_{1:t}$ can be obtained as

$$P(s_t = R|y_{1:t}) = \frac{P(y_t|s_t = R)\, P(s_t = R|y_{1:t-1})}{\sum_{j=E}^{R} P(y_t|s_t = j)\, P(s_t = j|y_{1:t-1})} \tag{7}$$

where $P(s_t = j|y_{1:t})\ \ j = E, R$ are referred to as filtered probabilities and $P(y_t|s_t = j)$ is the likelihood of the data in period t conditional on state j.

I estimate the model in equation (5) for all monthly indicators listed in Table (2), except the price indices and the monetary aggregates.[8] In particular, following Anas et al. (2008) I consider industrial production, the unemployment rate and new passenger car registrations. In addition, I include the sentiment indices in the dataset, i.e. the Economic Sentiment indicator and the indices for confidence in industry and retail sales, and the stock market index, since these could potentially provide even timelier recession signals than the aforementioned hard economic indicators. For quarterly real GDP growth I estimate the well-established modified version of equation (5) proposed by Hamilton (1989), which only allows for regime shifts in the mean μ_{s_t} but not in the coefficient on the lagged dependent variable or the variance.

I estimate the univariate Markov-Switching models for the selected indicators with Bayesian techniques as described in Kim and Nelson (1999). This involves setting up a Gibbs sampler that iteratively draws the states $S_{1:T}$, the probabilities p_{EE} and p_{RR} and the remaining unknown parameters $\{\mu_j, \psi_j, \sigma_j,\}$ for $j = E, R$ from their respective conditional posterior distributions using the filter proposed by Hamilton (1989) and the multi-move sampler suggested by Carter and Kohn (1994). A normal inverse Wishart prior, that is assumed to be symmetrical across the two states, is used for the coefficients and the variance in equation (5), while the probabilities p_{EE} and p_{RR} are assumed to a priori follow a beta distribution.

In this real-time application, I estimate the univariate Markov-switching models for each indicator m with data up to period T_m^*, i.e. the period for which the most recent observation for that indicator is available. The real-time recession probability for the current period $T = T_m^* + k_m$ can thus be obtained as

$$\pi_T^m = P(s_T = R|y_{1:T_m^*}) = P^{k_m}\left(P(s_{T_m^*} = R|y_{1:T_m^*})\, P(s_{T_m^*} = E|y_{1:T_m^*})\right)',$$

where k_m is the publication lag of indicator m and the (2 x 2) matrix P contains the estimated transition probabilities $p_{ij},\ \ i, j = E, R$.

4.3.2 Markov-Switching Linear Opinion Pool

The combination of forecasts from different sources is a very popular way of increasing the accuracy of point forecasts (see, for example, Bates and Granger, 1969; Kuzin et al., 2011; Schwarzmüller, 2015; Stock and Watson, 2003). As shown by Clements and Harvey (2011) and Ranjan and Gneiting (2010), among many others, the concept of forecast pooling can also be extended to probabilistic forecasts.

[8]For these variables, it is not intuitively clear that the assumption $\mu_E > \mu_R$ identifies a "high" state E and a "low" state R which correspond to phases of economic expansion and recession, respectively.

Following Anas et al. (2008), who construct a business cycle coincident indicator (BCCI) as a linear opinion pool of the probabilistic forecasts obtained with several univariate Markov-switching models to assess recession risks in the euro area, I implement a Markov-switching linear opinion pool with equal weights as

$$\pi_T^{pool} = M^{-1} \sum_{m=1}^{M} \pi_T^m. \tag{8}$$

While Anas et al. (2008) chose their pooling weights to minimize first and second order forecast errors, I opt for an equal-weight pool. The main reason for this is technical, because in this real-time application, where the realized business cycle phases are observed with a substantial delay, it would be very hard to compute meaningful pooling weights based on the recent forecast performance of the alternative forecast approaches. In addition, the limited sample size impedes the calculation of performance-based weights over a presample. Moreover, for point forecasts, equally weighted forecast pools have proven to be extremely competitive in comparison to pools with performance-based pooling weights (see, for example, Stock and Watson, 2004; Timmermann, 2006).

However, Ranjan and Gneiting (2010) show that for probabilistic forecasts, in general, the linear opinion approach is suboptimal, since it yields pools that are uncalibrated and lack sharpness. They propose to recalibrate the linear opinion pool by applying a beta transform which is given as

$$\pi_T^{pool,opt} = H_{\alpha,\beta}\left(M^{-1} \sum_{m=1}^{M} \pi_T^m \right), \tag{9}$$

where $H_{\alpha,\beta}$ is the cumulative distribution function of a beta density with parameters α and β for which $\alpha = \beta \geq 1$. I apply the beta transform in the robustness analysis in section (7) to assess ex post the degree to which the performance of the linear equal-weight pool is inferior to that of the optimal pool.

In total, I implement three linear equal-weight pools. The first combines the probabilistic forecasts of all considered univariate Markov-switching models, while the second combines only the real-time recession probabilities obtained with the models for the soft indicators, i.e. the two confidence indices in industry and retail sales, the Economic Sentiment indicator and the stock market index. Finally, in the spirit of the BCCI proposed by Anas et al. (2008), I implement a pool that aggregates the predictions of the univariate Markov-switching models for industrial production, the unemployment rate and new passenger car registrations.

4.4 A Google Trends Real-Time Recession Indicator for the Euro Area

Google Trends (`https://www.google.de/trends/`) provides real-time indices of the relative volume of Internet search queries for specific terms in a predefined geographic area starting from January 2004.[9] A growing body of literature has documented the usefulness of these data

[9] See Choi and Varian (2009b) for a description of the Google Trends interface and potential uses of the data.

to predict variables such as unemployment (Askitas and Zimmermann, 2009; Choi and Varian, 2009a), consumer demand and sales (Fantazzini and Toktamysova, 2015; Vosen and Schmidt, 2011; Yan and Labbé, 2013) as well as tourism flows (Concha and Galán, 2012) and influenza outbreaks Ginsberg et al. (2009).

As an alternative to the econometric real-time recession indices discussed above, I construct a euro area real-time recession indicator based on Google Trends data for the search query share of the word *recession*. In particular, the indicator is built as a population-weighted mean over the indices for the eleven largest euro area countries for which a query series is available. The list of countries includes Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Portugal and Spain.

The idea behind this approach is very similar to the *R-word index* introduced by *The Economist* magazine in the early 1990s, which tracks the number of newspaper articles that use the word *recession* in a given quarter. The *R-word index* has been found to be a reliable source of early signals for pending recessions in the US (Doms and Morin, 2004), Germany (Mayr and Grossarth-Maticek, 2008) and Switzerland (Iselin and Siliverstovs, 2013).

5 Evaluation of Probabilistic Recession Forecasts

I evaluate the real-time recession probabilities of the alternative approaches outlined in section (4) with formal scoring rules for the period ranging from January 2004 until December 2013, i.e. a total of 120 recession predictions are considered for the evaluation. In particular, the recession probability forecasts π_t are compared to a binary indicator variable bc_t that is equal to one for periods that were declared recessions by the CEPR and zero otherwise (see Table (1)).

The first scoring rule that I compute to assess the accuracy of the alternative approaches is the widely used quadratic probability score (QPS) which is given as

$$QPS = T^{-1} \sum_{t=1}^{T} \left(\pi_t - bc_t \right)^2 . \tag{10}$$

Gneiting et al. (2007) show that this score is proper, meaning that the forecaster has no incentive to state anything but his or her true beliefs. The QPS corresponds to the common notion of mean squared error loss that is typically used to evaluate point forecasts. That implies that the score explicitly accounts for the strength of false signals, meaning that a recession probability $\pi_t^1 = 0.8$ in a month where $bc_t = 0$ is considered to be worse than $\pi_t^2 = 0.6$.

The QPS simultaneously addresses the sharpness and calibration of the probabilistic forecasts π_t. It can be decomposed to make the performance in both dimensions visible. The negatively-oriented component that assesses the calibration of the probabilistic forecast is given as

$$CAL = T^{-1} \sum_{j=1}^{J} T_j^{-1} \left(\overline{\pi}_j - \overline{bc}_j \right)^2 , \tag{11}$$

while the positively-oriented sharpness component reads

$$SHARP = T^{-1} \sum_{j=1}^{J} T_j^{-1} \left(\overline{bc}_j - \overline{bc}_t \right)^2. \tag{12}$$

$\overline{\pi}_j \in [0,1]$ are $j = 1, ..., J$ discrete probability values used to define probability bins. T_j is the number of times π_t falls into bin j. $\overline{bc}_j$ is the respective empirical conditional event frequency and $\overline{bc}_t$ is the unconditional mean of bc_t (Ranjan and Gneiting, 2010). It holds that

$$QPS = CAL - SHARP + Var(bc_t). \tag{13}$$

I assess the statistical significance of the difference between the QPS scores for the alternative forecasting approaches with a version of the Diebold-Mariano test (Diebold and Mariano, 1995) that accounts for serial correlation of the forecast errors using Newey-West standard errors as proposed by Lopez (2001).

Lahiri and Wang (2013) survey a number of alternative methods that are well-suited to evaluating probabilistic forecasts for a decline in GDP. Unlike global measures of forecast quality such as the QPS, these measures explicitly take into account the ability of a forecasting approach to assess the odds for the occurrence of an event against its non-occurrence. This could be particularly important, for example, in the policy process when clear signals for the predicted occurrence or non-occurrence of an event have to be issued.

I apply two of the evaluation methods outlined in Lahiri and Wang (2013), namely the receiver operating characteristic (*ROC*) and the Peirce skill (*PS*) score. Both of these measures are based on (2x2) contingency tables which classify $\{\widehat{bc}_t\}_{t=1}^{T}$, the binary forecasts for the occurrence or non-occurrence of an event, into *Hits* ($\widehat{bc}_t = bc_t = 1$), *False Alarms* ($\widehat{bc}_t = 1, bc_t = 0$), *Misses* ($\widehat{bc}_t = 0, bc_t = 1$) and *Correct rejections* ($\widehat{bc}_t = bc_t = 0$) for a given period of observations $\{bc_t\}_{t=1}^{T}$. These binary event forecasts $\widehat{bc}_t$ can be obtained from the probabilistic forecasts π_t via a binary event classifier w, such that $\widehat{bc}_t = 1$ if $\pi_t > w$ and $\widehat{bc}_t = 0$ otherwise.

The *ROC* is calculated for a range of thresholds w and thus explicitly accounts for the role of the binary event classifier for the accuracy of the binary forecast signal. The *ROC* is commonly depicted as a curve of the rates of *Hits* against the corresponding rates of *False alarms* over a range of thresholds w for a given period of observations $\{bc_t\}_{t=1}^{T}$. Ideally, for high values of w, the rate of *Hits* should increase monotonically from zero to one as w decreases, while the rate of *False alarms* should remain constant at zero. For further decreases in w, the ideal *ROC* curve would indicate increasing *False alarm* rates but a constant *Hit* rate of one (see Figure (2)). By contrast, a *ROC* curve along the 45 degree line in the unit square indicates no discriminatory skill for the occurrence and non-occurrence of an event.

Alternatively, the *ROC* score can also be expressed as the area above the *ROC* curve. From the description of the ideal *ROC* curve, it is clear the *ROC* score $\in \{0,1\}$ and that it is zero for the ideal forecasting method which perfectly discriminates between the occurrence and non-occurrence of an event.

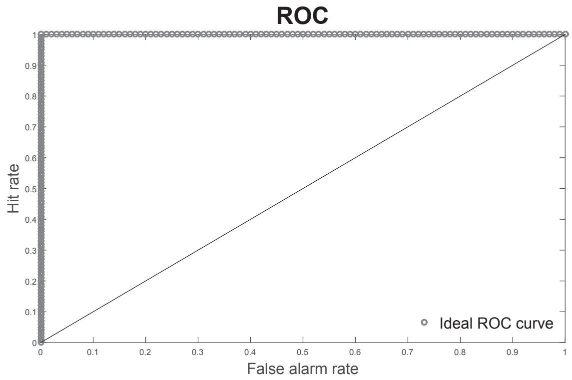


Figure 2: Ideal *ROC* curve.

Finally, the *PS* score is computed as the difference between the rate of *Hits* (H) and the rate of *False alarms* (F) for a given threshold w, i.e.

$$PS(w) = \frac{\sum_{t=1}^{T}(bc_t\widehat{bc}_t)}{\sum_{t=1}^{T} bc_t} - \frac{\sum_{t=1}^{T}\widehat{bc}_t - \sum_{t=1}^{T}(bc_t\widehat{bc}_t)}{T - \sum_{t=1}^{T} bc_t} = H - F. \tag{14}$$

For an ideal forecasting approach $PS(w) = 1$ or $PS(w) = -1$, whereby the latter value indicates that the binary signals are perfectly mislabeled. By contrast, $PS(w) = 0$ indicates no discriminatory skill at all. Following Lahiri and Wang (2013), I assess the statistical significance of the *PS* scores for the alternative forecasting approaches using the following standard error formula:

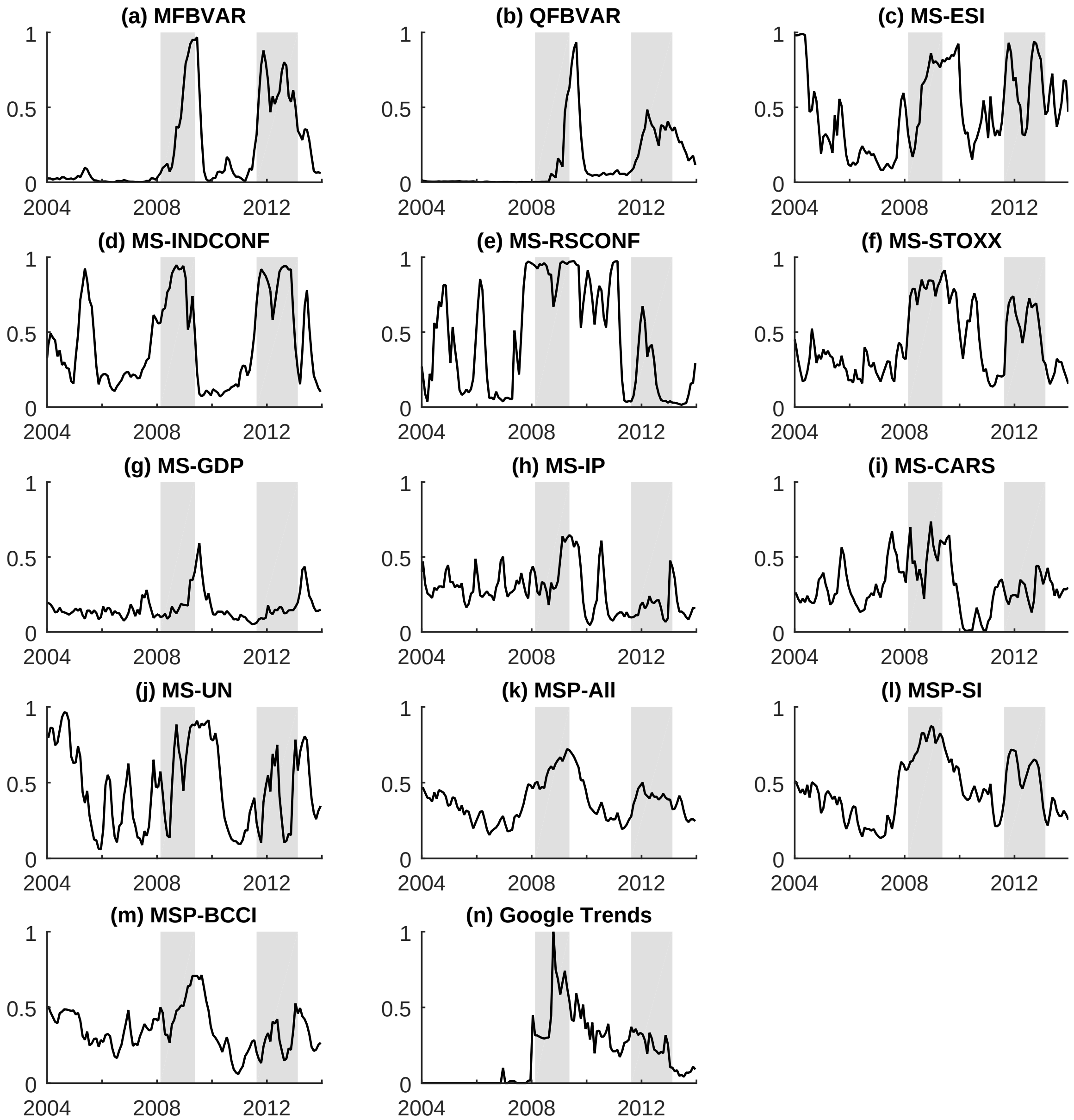
$$SE(w) = \sqrt{\frac{H(1-H)}{\sum_{t=1}^{T} bc_t} + \frac{F(1-F)}{T - \sum_{t=1}^{T} bc_t}} \tag{15}$$

6 Results

The monthly real-time recession signals obtained with the different methods described in section (4) are depicted in Figure (3). Since the original real-time recession probabilities obtained with the alternative approaches are very noisy, the real-time signals displayed are obtained as three-month weighted moving averages over the original probabilities.[10] The two recessions in the evaluation period from January 2004 until December 2013, as dated by the CEPR (see Table 1), are again indicated by the shaded areas in each of the panels.

In general, all the approaches shown in Figure (3), show increased real-time recession signals during the Great Recession and the recession related to the European debt crisis. However, there are considerable differences between the alternative approaches with respect to the timeliness of recession signals as well as the amount of *False alarms*, i.e. recession signals in non-recession periods. This is also reflected in Table (3), which contains the *QPS* for the alternative

[10] In particular, the real-time recession signals are obtained as $\tilde{\pi}_t = \frac{1}{6}\pi_{t-2} + \frac{2}{6}\pi_{t-2} + \frac{3}{6}\pi_t$, where π_t is the original recession probability for period t.



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas indicate euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. MFBVAR: mixed-frequency Bayesian vector autoregression, QFBVAR: quarterly BVAR. MS: univariate Markov-switching model for ESI: the Economic Sentiment indicator, INDCONF: index for confidence in industry, RSCONF: index for confidence in retail sales, STOXX: stock market indicator, GDP: real gross domestic product, IP: industrial production, CARS: new passenger car registrations, UN: unemployment rate. MSP: combination of probabilistic forecasts from univariate Markov-switching models for ALL: all univariate MS models, SI: the Economic Sentiment index, the confidence indices in industry and retail sale and the stock market index, BCCI: industrial production, the unemployment rate and new passenger car registrations.

Figure 3: Real-time recession signals for the euro area.

models in the first column as well as the results of the formal assessment carried out to measure the calibration (CAL) and sharpness ($SHARP$) of the probabilistic forecasts of the different approaches in the second and third column. Small values for QPS and CAL reflect a high overall level of accuracy and a good calibration, respectively, while high values for $SHARP$ indicate that the probabilistic forecasts of the alternative approaches are sharp.

	QPS	CAL	SHARP
MFBVAR	0.108	0.028	0.108
QFBVAR	0.211***	0.046	0.048
MS-ESI	0.231***	0.123	0.040
MS-INDCONF	0.110	0.074	0.114
MS-RSCONF	0.317***	0.161	0.005
MS-STOXX	0.148	0.075	0.087
MS-GDP	0.228***	0.033	0.006
MS-IP	0.227***	0.059	0.006
MS-CARS	0.188***	0.022	0.016
MS-UN	0.315***	0.173	0.009
MSP-All	0.168**	0.057	0.056
MSP-SI	0.154	0.100	0.090
MSP-BICC	0.215***	0.053	0.002
Google Trends	0.148	0.021	0.047

Notes: The real-time recession signals of the alternative approaches are evaluated over the sample from January 2004 until December 2013 using the CEPR business cycle chronology as a benchmark. QPS: quadratic probability score, CAL: calibration score, SHARP: sharpness score. ***(**,*) denote that the QPS is significantly different from the QPS of the MFBVAR at the 1% (5%,10%) level. For the model abbreviations see the notes to Figure (3).

Table 3: Evaluation of real-time recession probabilities, QPS.

Overall, the results presented in Table (3) suggest that MFBVAR performs best, as it achieves the lowest QPS among all the approaches considered here. However, the improvements of the MFBVAR over the univariate Markov-switching models for industry confidence, the stock market, the pool of models for the sentiment indices and the Google Trends real-time recession index are not statistically significant, as indicated by the results of the respective pairwise Diebold-Mariano tests.

The Markov-switching models for industry confidence achieves a QPS that is only slightly higher than that of the MFBVAR, although its real-time recession signals are less well-calibrated. This is confirmed in Figure (3), which shows that the small number of *False alarms* produced by the MFBVAR in panel (a) are considerably less pronounced than those of the model for the industry confidence index in panel (d). On the other hand, the latter model performs better than the MFBVAR in terms of sharpness, reflecting the fact that the real-time recession signals of the MFBVAR at the onset of the Great Recession are only very muted.

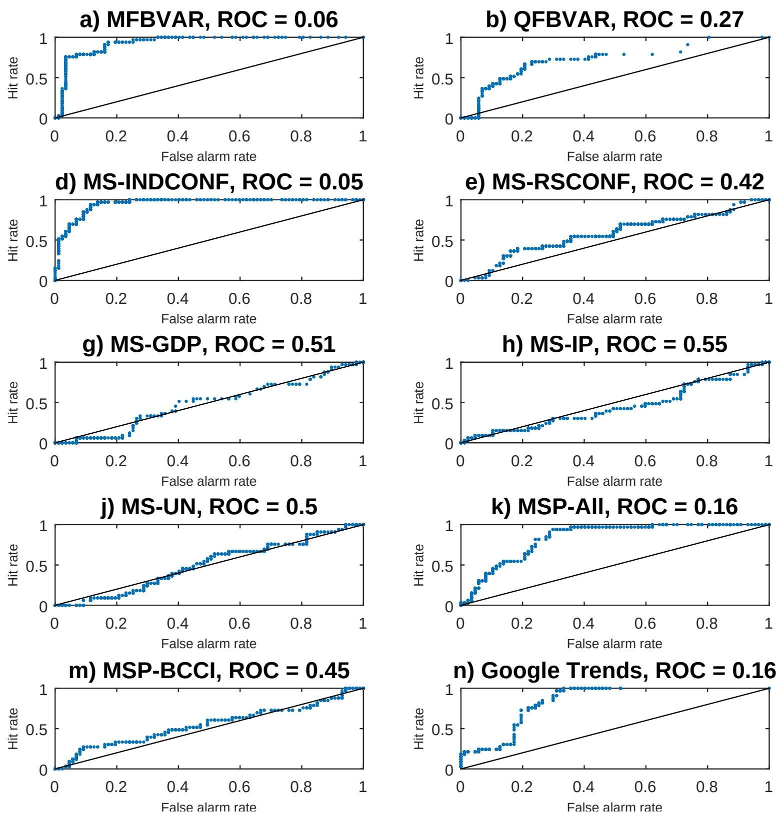
The univariate Markov-switching models for the stock market index, the pool of all sentiment indices and the Google Trends real-time recession indicator perform more or less equally well in terms of QPS. However, for the former two, this is due to the high sharpness of their forecasts, while the Google Trends real-time recession signals are better calibrated. Again, this is confirmed in Figure (3), which shows that the real-time recession signals obtained with the models for the stock market index in panel (f) and the sentiment pool in panel (k) are concentrated at the confident values of zero and one. On the other hand, both models issue many *False alarms*, pushing down their relative performance in terms of calibration. By contrast, the Google Trends indicator depicted in panel (n) is almost always equal to zero in the non-recession period prior to 2008 and also very low after the end of the second recession in the sample. This improves the calibration score of the index considerably. Its moderate performance in terms of sharpness can be explained by the fact that, apart from two drastic increases in euro area Internet users' interest in the word *recession* in early 2008 and in September 2008, the signals were mostly not very clear-cut, especially for the period between the two recessions in the sample. One reason for the latter observation could be the general high level of uncertainty in that period about the sustainability of early signs of economic relief and fears of a double-dip recession (Camacho et al., 2014).

The results for the univariate Markov-switching models for real GDP growth, industrial production and new passenger car registrations (panels (g) - (i) in Figure (3)) illustrate that a well-calibrated probabilistic forecast can be of minimal use in practice if it lacks sharpness. The QPS for these models is very high. To a lesser extent, this also applies to the QFBVAR in panel (b), the pool of all Markov-switching models in panel (k) and the BCCI pool considered in Anas et al. (2008) in panel (m). What all these models have in common is that they deliver real-time recession signals that are not very clear-cut and are, for the most part, also heavily delayed. However, given that these models are estimated with the series that have the highest publication lag (see Table (2)) this result is not particularly surprising.

By contrast, the poor performance of the two sentiment indices, namely the Economic Sentiment indicator and the confidence index in retail sales, might come as something of a surprise. Both models apparently not only lack sharpness but are also very poorly calibrated. Indeed, from panel (c) and (e) in Figure (3) it can be seen that these two sentiment indices deliver many pronounced *False alarms*. One possible reason for this could be that these sentiment indices might not only be driven by hard economic fundamentals, but also by other factors. These could possibly be unrelated contagious waves of optimism and pessimism which are often referred to as animal spirits or noise shocks (Akerlof and Shiller, 2008; De Grauwe, 2011).

Finally, the univariate Markov-switching model for the unemployment rate achieves by far the highest QPS of all models considered here. The pattern of the model's real-time recession signals depicted in panel (j) suggests that the unemployment rate is likely to increase only with a certain lag after the beginning of a recession. In fact, this confirms that this variable is typically regarded as a lagging rather than a contemporaneous or even leading indicator for the state of the economy. In addition, the model also clearly reflects the steady increase in euro area unemployment until early 2005, which was not accompanied by a recession. On closer inspection, panel (j) and panel (c) of Figure (3) reveal some similarities between the real-time recession signals obtained with the unemployment rate and those delivered by the model for the Economic Sentiment indicator. This could suggest that the latter is driven by news about the former. However, much more research would be needed to establish a solid causal link here, of course.

Figure (4) depicts the *ROC* curves for the alternative approaches as well as the respective *ROC* scores, i.e. the areas above the *ROC* curve in the unit square. It is plain to see that the *ROC* curves for the MFBVAR in panel (a) and for the model used for the confidence index in industry in panel (d) are far away from the 45-degree line and that they achieve the lowest *ROC* scores. This supports the notion that the discriminatory skill of the two approaches is very high overall, independent of the selected binary event classifier w. The *ROC* scores for the model used for the stock market index in panel (f), the Markov-switching pool of sentiment indices in panel (l) and, to a slightly lesser extent, the pool of all univariate models in panel (k) and the Google Trends indicator in panel (n) are moderately higher but still considerably lower than those for the remaining models. In particular, the models for the hard indicators depicted in panels (g) - (j) and the BCCI pool considered by Anas et al. (2008) in panel (m) produce very flat *ROC* curves, which indicates that the models' real-time recession signals are unable to discriminate between recession and non-recession periods.



Notes: The curves displayed show the rate of *Hits* and the corresponding rate of *False alarms* for varying thresholds used to transform the monthly real-time recession probabilities displayed in Figure (3) into binary signals for the occurrence and non-occurrence of a recession. The evaluation sample ranges from January 2004 until December 2013. ROC: receiver operating characteristic (area above the depicted curve). A small ROC denotes high discriminatory skill. For the model abbreviations see the notes to Figure (3).

Figure 4: *ROC* curves and scores.

	PS				
$w =$	0.1	0.2	0.3	0.4	0.5
MFBVAR	0.75***	0.66***	0.71***	0.66***	0.60***
QFBVAR	0.43***	0.35***	0.32***	0.09	-0.03
MS-ESI	0.03*	0.22***	0.28***	0.26**	0.39***
MS-INDCONF	0.09***	0.38***	0.67***	0.73***	0.82***
MS-RSCONF	0.06	0.11	0.18	0.11	0.10
MS-STOXX	0.00	0.21***	0.48***	0.70***	0.72***
MS-GDP	-0.04	-0.12	-0.02	-0.07	-0.02
MS-IP	0.05**	0.15*	0.19*	0.08	0.04
MS-CARS	0.13***	0.20***	0.25***	0.24***	0.09
MS-UN	0.06**	-0.00	0.05	0.11	0.03
MSP-All	0.00	0.10	0.39***	0.58***	0.28***
MSP-SI	0.00	0.15	0.38***	0.48*	0.68***
MSP-BICC	0.05	-0.02	0.03	0.08	0.13*
Google Trends	0.64***	0.58***	0.41***	0.15*	0.17**

Notes: The Peirce skill score (PS) is calculated as the difference between the rate of *Hits* and the corresponding rate of *False alarms* for the binary event classifier w. The evaluation sample ranges from January 2004 until December 2013. For the model abbreviations see the notes to Figure (3).

Table 4: Evaluation of real-time recession probabilities, PS.

The entries in Table (4) reveal the usefulness of the alternative recession signals for selected binary event classifiers w and could give valuable guidance to practitioners. For example, the MFBVAR achieves the highest significant PS score for a very small threshold, while the other high-performing models, i.e. those for the indices of the stock market and industry confidence and the pool of models for the sentiment indices, perform best for the common threshold of $w = 0.5$. By contrast, the PS score for the Markov-switching models for the confidence index in retail sales, real GDP, industrial production and the unemployment rate is not statistically different from zero, indicating that these models do not have any discriminatory skill for the real-time detection of recession periods. For the remaining approaches, the threshold for which they are most useful varies. For the QFBVAR and the Google Trends indicator, a low threshold works best, while the Markov-switching pools and the univariate models for new passenger car registrations yield the highest PS score for an intermediate threshold. Overall, these results illustrate that the absolute values of the real-time recession signals issued by the alternative approaches should be interpreted very carefully. In particular, signals obtained with different approaches which are equal in absolute value cannot necessarily be interpreted as being equally strong.

a) MFBVAR, $w = 0.1$		
	$bc_t = 1$	$bc_t = 0$
$\widehat{bc}_t = 1$	30	14
$\widehat{bc}_t = 0$	3	73
b) INDCONF, $w = 0.5$		
	$bc_t = 1$	$bc_t = 0$
$\widehat{bc}_t = 1$	32	13
$\widehat{bc}_t = 0$	1	74

Notes: The entries display the number of *Hits* ($\widehat{bc}_t = bc_t = 1$), *False Alarms* ($\widehat{bc}_t = 1$, $bc_t = 0$), *Misses* ($\widehat{bc}_t = 0$, $bc_t = 1$) and *Correct rejections* ($\widehat{bc}_t = bc_t = 0$) over the period from January 2004 until December 2013. $bc_t = 1$ ($bc_t = 0$) denotes a recession (expansion) month. $\widehat{bc}_t = 1$ ($\widehat{bc}_t = 0$) denotes a recession (expansion) forecast. For the model abbreviations see the notes to Figure (3).

Table 5: Contingency table, MFBVAR and INDCONF.

According to the entries in Table (4) the MFBVAR achieves a maximum *PS* score that is slightly lower than that of the univariate Markov-switching model for the industry confidence index. To understand the significance of this difference, it might be warranted to take a closer look at the contingency table for both models. Table (5) shows the contingency table for the threshold value w that maximizes the *PS* scores of the two models. The columns contain the number of months in which a recession did and did not occur in the period from January 2004 until December 2013, and the rows display the number of times these events were forecast. Out of a total of 33 recession months in the sample, the MFBVAR misses three months, while the industry confidence index model misses only one. Moreover, for the 87 non-recession months the MFBVAR issues one more *False Alarm* than the model for the industry confidence index. Given the high uncertainty surrounding any attempt to precisely determine the start and end months of a recession, these differences seem to be rather small.

It should nonetheless be noted that, depending on the purpose of the forecast and the forecaster's loss function, missed recessions and *False Alarms* could actually be very costly and even small differences between alternative approaches might be very relevant in practice. It is also important to be aware of the characteristics of the alternative evaluation procedures. For example, all formal measures used above treat *Misses* and *False Alarms* symmetrically, which might be inappropriate when the economic costs of the two differ (see Knedlik, 2014, for an application with asymmetric weights). As pointed out by Lahiri and Wang (2013), for known economic costs of these two types of errors, decision theoretic frameworks could be used to derive binary real-time recession signals. However, this goes beyond the scope of this paper.

In summary, the results of the formal evaluation presented in this section indicate that the MFBVAR and the Markov-switching model for the confidence index in industry deliver the most accurate real-time recession signals in terms of calibration, sharpness and discriminatory skill, and that they perform more or less equally well.

7 Robustness

In the following robustness analysis, I first investigate whether the performance of the MFB-VAR described in section (6) is robust with respect to two aspects, namely the variables that are included in the model and the recession definition used to compute the MFBVAR real-time recession probabilities. Second, I assess the extent to which the performance of the linear opinion pools can be improved by applying the beta transform proposed by Ranjan and Gneiting (2010) and described in section (4.3.2).

Regarding the variables included in the estimation of the MFBVAR, the results of section (6) suggest that the sentiment indices could be crucially important for the accuracy of the model's real-time recession signals. To verify this conjecture, I estimate a version of the model that excludes the Economic Sentiment index and the confidence indices in construction and retail sales ($MFBVAR(9)$) from the list of variables in Table (2). Additionally, I assess the performance of a version of the MFBVAR that includes only the most useful sentiment index, i.e. the confidence index for the industry sector, and quarterly real GDP ($MFBVAR(2)$).

To assess the robustness of the performance of the MFBVAR with respect to the recession definition used to extract real-time recession probabilities from the model's predictive distribution, I consider the following alternatives. First, I assess the performance of MFBVAR real-time recession signals based on weak real GDP growth rates below 0.1%, rather than below zero, over a sequence of at least two consecutive quarters ($MFBVAR^{\text{slow growth}}$). This might increase the timeliness of the model's real-time recession signals, since at the onset of most recessions, the GDP decline is often quite muted (see panel (b) of Figure (1)). Second, I consider a multivariate approach that is closer to the recession definition used by the CEPR (Artis et al., 2003), see also section (2)), and which, among others, also requires monitoring employment conditions and the state of the industrial sector in the euro area. In particular, I use the MFBVAR's joint predictive distribution of real GDP, industrial production, the unemployment rate, the Economic Sentiment index, the confidence index in industry and the stock market index to compute the odds that several of these variables simultaneously signal a deterioration in economic conditions. I include the soft indicators so as not to run the risk of the recession signals being delayed. However, since my previous results suggest that the soft indicators are prone to delivering *False alarms*, I define this alternative recession criterion as the odds that four out of the six above-listed indicators deteriorate repeatedly for at least three consecutive months ($MFBVAR^{4/6}$). As a final alternative, I use the MFBVAR estimate for latent monthly real GDP to assess recession risks in real-time, which I define as the odds that real GDP growth in the current month is part of a three-month sequence of negative growth rates ($MFBVAR^{\text{monthly GDP}}$).

Table (6) presents the results of my robustness analysis for the performance of the MFB-VAR, while the corresponding real-time recession signals are depicted in figure (6) in appendix A 3. The first row contains the results of the benchmark model version as presented in section (6), while the results for the model versions that include alternative sets of indicators are presented in the middle part of the table. Finally, the last rows contain the results for the benchmark model including all variables listed in Table (2), but here, the aforementioned alternative recession definitions were used for the computation of the real-time recession signals.

	QPS	*ROC*	PS*
MFBVAR	0.11	0.06	0.75 ($w = 0.1$)
MFBVAR(2)	0.11	0.10	0.67 ($w = 0.3$)
MFBVAR(9)	0.15***	0.13	0.59 ($w = 0.2$)
MFBVAR$^{\text{slow growth}}$	0.09**	0.06	0.72 ($w = 0.5$)
MFBVAR$^{4/6}$	0.13	0.04	0.67 ($w = 0.1$)
MFBVAR$^{\text{monthly GDP}}$	0.12	0.06	0.69 ($w = 0.2$)

Notes: The evaluation sample ranges from January 2004 until December 2013. QPS: quadratic probability score, ROC: receiver operating characteristic, PS*: maximum Peirce skill score obtained with the binary event classifier w.***(**,*) denote that the QPS is significantly different from the QPS for the MFBVAR at the 1% (5%,10%) level. See the text for the model abbreviations.

Table 6: Robustness of MFBVAR performance.

According to the entries in Table (6), the performance of the model clearly deteriorates when all sentiment indices are excluded. The MFBVAR(9) achieves a significantly larger QPS than the benchmark, while the ROC score and the maximum PS score both decrease. The most important variable, however, turns out to be the confidence index in industry. The MFBVAR(2) that only includes this index in addition to real GDP achieves the same QPS as the benchmark model, and only slightly higher ROC and maximum PS scores.[11]

Regarding the alternative conditions for defining real-time recession risks, only the version based on slow growth rather than negative growth rates can significantly improve upon the benchmark in terms of the QPS, while the approaches based on the joint deterioration of four monthly indicators or the estimate for latent monthly real GDP perform just as well. Moreover, the differences between the three alternatives in terms of ROC scores and maximum PS scores are very small. Hence, the performance of the MFBVAR real-time recession signals turns out to be quite robust with respect to the exact definition of real time recession risks.

Finally, it remains to be investigated whether the performance of the Markov-switching linear equal-weight pools can be improved upon by applying the beta transformation proposed by Ranjan and Gneiting (2010) and described in section (4.3.2). The upper part of Table (7) repeats the results from section (6) for the linear equal-weight pools, while the lower part of Table (7) presents the results for the ex post best-performing beta transformed pools in terms of the QPS score. These were obtained through a grid search over alternative values for α, the parameter of the beta distribution. The real-time recession signals of the original and the optimally transformed are depicted in figure (6) in appendix A 4.

As it turns out, the optimal beta transform can improve the performance of the alternative pools relative to the untransformed linear equal-weight pool, especially in terms of calibration. However, the overall gains in accuracy are rather small.

[11] An evaluation of the point and density forecasts of the different model versions can be found in appendix A 1.

	QPS	CAL	SHARP	ROC	PS*
MSP-All	0.168**	0.057	0.056	0.16	0.58 ($w = 0.4$)
MSP-SI	0.154	0.100	0.090	0.09	0.68 ($w = 0.5$)
MSP-BICC	0.215***	0.053	0.002	0.45	0.13 ($w = 0.5$)
MSP-All, $\alpha* = 3.4$	0.151*	0.028	0.060	0.16	0.57 ($w = 0.1$)
MSP-SI, $\alpha* = 6.1$	0.128	0.066	0.093	0.10	0.68 ($w = 0.5$)
MSP-BICC, $\alpha* = 1$	0.215***	0.053	0.002	0.45	0.13 ($w = 0.5$)

Notes: See the notes to Table (6). CAL: calibration score, SHARP: sharpness score. MSP: combination of probabilistic forecasts fromunivariate Markov-switching models for ALL: all univariate MS models, SI: the Economic Sentiment index, the confidence indices in industry and retail sales and the stock market index, BCCI: industrial production, the unemployment rate and new passenger car registrations. Upper part of the table: linear pools, lower part: optimal beta transform.

Table 7: Robustness of the performance of the linear opinion pools.

Moreover, in comparison to the MFBVAR, the optimal pool that includes all univariate Markov-switching models and the optimal BCCI pool in the spirit of Anas et al. (2008) perform significantly worse still, while for the optimal pool of models for the sentiment indices, there is no statistically significant difference.

8 Conclusion

The evidence presented in this paper shows that the predictive distribution of the back-, now- and forecasts obtained with a linear mixed-frequency Bayesian VAR (MFBVAR) can be used to extract very accurate monthly real-time recession signals for the euro area. Evaluated over the period from January 2004 until December 2013, the probabilistic real-time recession forecasts of the MFBVAR outperform those obtained with the univariate regime-switching models for a number of hard and soft monthly economic indicators, their linear combinations and a real-time recession index obtained with Google Trends data as measured by the quadratic probability score, the receiver operating characteristic and the Peirce skill score. Only the univariate Markov-switching model for the confidence index in industry delivers real-time recession signals that are more or less as accurate as those of the MFBVAR.

The real-time recession signals obtained with the remaining soft indicators, namely the Economic sentiment index and the indicator for confidence in retail sales, are very poorly calibrated and yield a high number of recession signals in non-recession periods. This could suggest that these variables are possibly driven by more than just economic fundamentals. The hard economic indices considered here, i.e. industrial production and quarterly real GDP growth, perform particularly poorly in terms of sharpness and thus have no discriminatory skill to separate recessions from periods of economic expansion in real-time. The reason for this is most likely the long publication lag of the respective data. The Google Trends real-time recession index, which is obtained as a population-weighted mean of the query shares for the word ”recession” in the eleven largest euro area countries, accurately signals the beginning of the Great Recession in 2008-09 and the end of the recession related to the European debt crisis in 2011-13. However, in between these two recession periods, its signals are not very clear-cut, possibly on account of the uncertainty prevailing at that time over the occurrence of a double-dip recession.

The robustness analysis indicates that the inclusion of the confidence index in industry is crucial for the good performance of the MFBVAR. Moreover, the results show that considering a multivariate recession definition that requires the joint monitoring of several soft and hard economic indicators does not significantly increase the accuracy of the MFBVAR's real-time recession signals compared to the benchmark case, where only the evolution of real quarterly GDP growth is assessed.

Further, the robustness analysis provides evidence that the performance of the linear opinion pools of the probabilistic forecasts of the various univariate regime-switching models can be increased if an optimal beta transformation as suggested by Ranjan and Gneiting (2010) is applied. However, even the ex-post optimized pools are not more accurate than the MFBVAR or the univariate Markov-Switching model for the confidence index in industry.

Finally, the findings illustrate that the absolute values of the real-time recession signals issued by the alternative approaches should be interpreted very carefully. The size of the optimal binary event classifier used to translate the probabilistic recession forecasts into binary signals for the occurrence or non-occurrence of a recession varies considerably. In particular, the widely used threshold of 0.5 turns out to be suboptimal in many cases.

References

Aastveit, K. A., A. S. Jore, and F. Ravazzolo (2014). Forecasting recessions in real time. *Norges Bank Working Papers 02/2014*.

Akerlof, G. and R. Shiller (2008). *Animal Spirits*. Princeton University Press.

Anas, J., M. Billio, L. Ferrara, and G. L. Mazzi (2008). A system for dating and detecting turning points in the euro area. *The Manchester School 76*, 549 – 577.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011). Short-term forecasts of euro area GDP growth. *The Econometrics Journal 14*(1), 25 – 44.

Artis, M., F. Canova, J. Gali, F. Giavazzi, R. Portes, L. Reichlin, H. Uhlig, and P. Weil (2003). Business Cycle Dating Committee of the Centre for Economic Policy Research, Findings 22 September 2003.

Askitas, N. and K. Zimmermann (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly 55*, 107–120.

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1 – 27.

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447 – 460.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Banbura, M. and G. Rünstler (2011). A look into the factormodel black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting 27*(2), 333 – 346.

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

Bellgo, C. and L. Ferrara (2009). Forecasting euro-area recessions using time-varying binary response models for financial variables. *Banque de France Working Paper 259*.

Billio, M., R. Casarina, F. Ravazzolo, and H. K. van Dijk (2012). Combination schemes for turning point predictions. *The Quarterly Review of Economics and Finance 52*, 402 – 412.

Boysen-Hogrefe, J., N. Jannsen, M. Plödt, and T. Schwarzmüller (2015). An empirical evaluation of macroeconomic surveillance in the european union. *Kiel Working Papers, No. 2014*.

Bry, G. and C. Boschan (1971). Cyclical analysis of time series: Selected procedures and computer programs. *NBER Technical Paper 20*.

Camacho, M., G. P. Quiros, and P. Poncela (2014). Green shoots and double dips in the euro area: A real time measure. *International Journal of Forecasting 30*(3), 520 – 535.

Carter, C. and P. Kohn (1994). On Gibbs Sampling for State Space Models. *Biometrica 81*, 541 – 533.

Chauvet, M. and J. Piger (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics 26*(1), 42 – 49.

Choi, H. and H. Varian (2009a). Predicting initial claims for unemployment benefits. *Economics Research Group, Google.*.

Choi, H. and H. Varian (2009b). Predicting the present with google trends. *Economics Research Group, Google.*.

Clements, M. and D. Harvey (2011). Combining probability forecasts. *International Journal of Forecasting 27*, 208–223.

Concha, A. and E. Galán (2012). Tracking the future on the web: construction of leading indicators using internet searches. *Documentos Ocasionales N. 1203, Banco d'Espania*.

De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory 47*(2-3), 423–457.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*, 253–263.

Doms, M. and N. Morin (2004). Consumer sentiment, the economy, and the news media. *FRBSF Working Paper 2004-09*.

Dovern, J. and F. Huber (2015). Global prediction of recessions. *Economics Letters 133*, 81–84.

ECB (2008). Short-term forecast of economic activity in the euro area. *ECB Monthly Bulletin April*, 69 – 74.

Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model (AWM) for the euro area. *ECB working paper No. 42*.

Fantazzini, D. and Z. Toktamysova (2015). Forecasting german car sales using google data and multivariate models. *MPRA Paper No. 67110*.

Foroni, C. and M. Marcellino (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. *International Journal of Forecasting 30*(3), 554 – 568.

Geweke, J. and C. Whiteman (2006). *Handbook of Economic Forecasting*, Volume 1, Chapter 1 - Bayesian Forecasting, pp. 3–80. Elsevier Publications.

Giannone, D., J. Henry, M. Lalik, and M. Modugno (2010). An area-wide real-time database for the euro area. *ECB Working Paper Series No. 1145*.

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.

Ginsberg, J., M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant (2009). Detecting influenza epidemics using search engine query data. *Nature 457*, 1012–1014.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series: B 69*(2), 243 – 268.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357 – 384.

Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting 27*(4), 1006 – 1026.

Harding, D. and A. Pagan (2002). Dissecting the cycle: A methodological investigation. *Journal of Monetary Economics 49*, 365–381.

Iselin, D. and B. Siliverstovs (2013). The R-word index for switzerland. *Applied Economics Letters 20*(11), 1032–1035.

Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics 12(2)*, 99–132.

Karlsson, S. (2013). *Handbook of Economic Forecasting*, Volume 2, Chapter 15 - Forecasting with Bayesian Vector Autoregression, pp. 791–897. Elsevier Publications.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. MIT Press.

Knedlik, T. (2014). The impact of preferences on early warning systems - the case of the european commission's scoreboard. *European Journal of Political Economy 34*(C), 157–166.

Kuzin, V., M. Marcellino, and C. Schumacher (2011). Midas vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting 27*(2), 529 – 542.

Lahiri, K. and J. G. Wang (2013). Evaluating probability forecasts for GDP declines using alternative methodologies. *International Journal of Forecasting 29*, 175–190.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics 4*(1), 25–38.

Lo, M. C. and J. M. Piger (2005). Is the response of output to monetary policy asymmetric? Evidence from a regime-switching coefficients model. *Journal of Money, Credit, and Banking 37*(5), 865 – 886.

Lopez, J. A. (2001). Evaluating the predictive accuracy of volatility models. *Journal of Forecasting 20*(2), 87–109.

Marcellino, M. and C. Schumacher (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics 72*(4), 518 – 550.

Mayr, J. and J. Grossarth-Maticek (2008). Medienberichte als Konjunkturindikator. *ifo Schnelldienst 07*, 17–29.

Nalewaik, J. J. (2012). Estimating probabilities of recession in real time using GDP and GDI. *Journal of Money, Credit and Banking 44*(1), 235 – 253.

Österholm, P. (2012). The limited usefulness of macroeconomic bayesian vars when estimating the probability of a U.S. recession. *Journal of Macroeconomics 34*(1), 76 – 86.

Ranjan, R. and T. Gneiting (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(1), 71–91.

Reinhart, C. M. and G. L. Kaminsky (1999). The twin crises: The causes of banking and balance-of-payments problems. *American Economic Review 89*(3), 473–500.

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business and Economic Statistics, (forthcoming)*.

Schreiber, S. (2014). Anticipating business-cycle turning points in real time using density forecasts from a VAR. *Free University Berlin, School of Business & Economics Discussion Papers 2*.

Schumacher, C. (2014). MIDAS and bridge equations. *Deutsche Bundesbank Discussion Paper 26*.

Schumacher, C. and J. Breitung (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting 24*, 368 – 398.

Schwarzmüller, T. (2015). Model pooling and changes in the informational content of predictors: an empirical investigation for the euro area. *Kiel Working Papers 1982*.

Sims, C. and T. Zha (1998). Bayesian methods for dynamic multivariate models. *International Economic Review 39*(4), 949–968.

Stock, J. H. and M. W. Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23*(6), 405–430.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, pp. 135–196. Amsterdam: North Holland.

Vosen, S. and T. Schmidt (2011). Forecasting private consumption: survey-based indicators vs. google trends. *Journal of Forecasting 30*(6), 565–578.

Yan, C.-S. and F. Labbé (2013). Nowcasting with google trends in an emerging market. *Journal of Forecasting 32*, 289 – 298.

Appendix

A 1: Evaluation of MFBVAR and QFBVAR GDP Growth Goint and Density Forecasts

The following reports the results of the evaluation of the real GDP growth predictions provided by the MFBVAR and the QFBVAR. The evaluation period ranges from the first quarter of 2004 until the fourth quarter of 2013. To compute the different evaluation measures, the first release of euro area GDP growth is used to capture the real-time environment at the time the forecast was made. In all tables, $h^* = 1$ refers to the one-quarter-ahead forecast, while $h^* = 0$ and $h^* = -1$ denote the nowcast and the backcast, respectively. IS denotes the information set, i.e. the month of the current quarter, in which the back-, now- and forecast is computed.

The accuracy of the point forecasts is assessed with the mean squared forecast error, which is given as

$$MSFE = \frac{1}{T^*} \sum_{t=1}^{T^*} (y_t - \hat{y}_t)^2, \tag{16}$$

where y_t is the realized value of variable y in period t and $\hat{y}_t$ denotes the respective point forecast.

The accuracy of the density forecasts is assessed with several measures. The first is the logarithmic score, which is given as

$$LS = \frac{1}{T^*} \sum_{t=1}^{T^*} -log\big(F(y_t)\big),$$

where $F(.)$ denotes the predictive distribution.

The continuous ranked probability score (CRPS) is given as the average over

$$CRPS_t = \int_{-\infty}^{\infty} \big\{F(u) - \mathcal{I}(u \geq y_t)\big\}^2 du$$

or

$$CRPS_t = E_F|Y - y_t| - \frac{1}{2} E_F|Y - Y'|,$$

where E_P denotes the respective expectations operator and Y and Y' are random draws from the models' predictive cumulated distribution $F(.)$ (see, for example, Gneiting et al., 2007).

Note that all evaluation measures reported in the following are negatively oriented, i.e. the smaller the score, the better.

IS	1	2	3	1	2	3	1	2
		$h^* = 1$			$h^* = 0$		$h^* = -1$	
MFBVAR	0.51	0.32	0.36	0.26	0.25	0.19	0.16	0.12
MFBVAR(2)	0.45	0.37	0.34	0.30	0.32	0.24	0.24	0.28
MFBVAR(9)	0.48	0.34	0.41	0.34	0.27	0.26	0.19	0.14
QFBVAR	0.45	0.45	0.43	0.42	0.42	0.28	0.29	0.29

Table 8: RMSFE.

IS	1	2	3	1	2	3	1	2
		$h^* = 1$			$h^* = 0$		$h^* = -1$	
MFBVAR	1.26	1.14	1.13	1.08	1.07	1.03	1.00	0.97
MFBVAR(2)	1.22	1.16	1.13	1.11	1.12	1.07	1.05	1.06
MFBVAR(9)	1.26	1.16	1.16	1.12	1.08	1.05	1.01	0.98
QFBVAR	1.17	1.18	1.16	1.16	1.16	1.09	1.06	1.06

Table 9: Log score.

IS	1	2	3	1	2	3	1	2
		$h^* = 1$			$h^* = 0$		$h^* = -1$	
MFBVAR	0.35	0.28	0.29	0.25	0.24	0.22	0.20	0.18
MFBVAR(2)	0.33	0.29	0.28	0.27	0.28	0.25	0.25	0.26
MFBVAR(9)	0.34	0.28	0.31	0.27	0.24	0.23	0.21	0.19
QFBVAR	0.30	0.30	0.30	0.30	0.30	0.26	0.26	0.26

Table 10: CRPS.

A 2: MFBVAR Prior Specification

For the parameters of the VAR $A_1, ..., A_p, C$ and Σ, I implement a normal inverse Wishart prior that retains the main principles of the widely used Minnesota prior (Kadiyala and Karlsson, 1997; Litterman, 1986). This prior implies that $A_1, ..., A_p$ are assumed to be a priori independently and normally distributed, while with respect to the constant C the prior is assumed to be diffuse. The residual covariance matrix Σ is assumed to a priori follow an inverse Wishart distribution with scale matrix $\bar{S}$ and degrees of freedom $\bar{\alpha}$.

One of the main principles of the Minnesota prior is to center each equation of the VAR around a random walk with drift. Thus, the prior mean for $A_1, ..., A_p$ is specified as:

$$E[(A_\ell)_{ij}] = \begin{cases} 1 & \text{for} \quad i = j, \ell = 1 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Moreover, the prior also incorporates the belief that more recent lags of a variable should provide more reliable information for the estimation than less recent lags. The zero coefficient prior on more recent lags is therefore not imposed as tightly as on less recent lags. This is captured by specifying the prior variance as

$$Var[(A_\ell)_{ij}] = \begin{cases} \frac{\lambda^2}{\ell^2} & \text{for} \quad i = j \\ \frac{\lambda^2 \sigma_i^2}{\ell^2 \sigma_j^2} & \text{otherwise,} \end{cases} \tag{18}$$

where $\ell = 1, ..., p$ is the lag length, $\lambda = 0.2$ is a hyperparameter governing the importance of the prior beliefs relative to the data, and σ_i / σ_j is a scale parameter adjusting the prior for the different scale and variability of the data. For the implementation, σ_i is set equal to the standard deviation of the residuals of a simple univariate autoregression for each variable.

Following Schorfheide and Song (2015) I augment the prior outlined above to constrain the sum of coefficients of the VAR (Sims and Zha, 1998) as well as to incorporate the belief that the variables in the VAR follow a common stochastic trend.

I implement the prior outlined above using dummy variables Y^D and X^D, which are given as

$$Y^D = \begin{pmatrix} \frac{\text{diag}(\sigma_1 ... \sigma_n)}{\lambda} \\ 0_{n(p-1)xn} \\ \text{diag}(\sigma_1 ... \sigma_n) \\ 0_{1xn} \\ \text{diag}(\mu_1 ... \mu_n)\gamma \\ \eta\mu_1 ... \eta\mu_n \end{pmatrix}, X^D = \begin{pmatrix} \frac{\text{diag}(1...p) \otimes \text{diag}(\sigma_1 ... \sigma_n)}{\lambda} & 0_{npx1} \\ 0_{nxnp} & 0_{nx1} \\ 0_{1xnp} & c \\ P \otimes \text{diag}(\mu_1 ... \mu_n)\gamma & 0_{nx1} \\ (P' \otimes \text{diag}(\mu_1 ... \mu_n)\eta)' & \eta \end{pmatrix}, \tag{19}$$

where P is a (1xp) matrix of ones, $c = 10^{-4}$ reflects the diffuse prior for the constant C, $\mu_1, ..., \mu_n$ are the variable means and $\gamma = 10^3\lambda$ and $\eta = \lambda$ govern the tightness of the sum of coefficients constraint and the common stochastic trend prior, respectively.

With these dummy variables, the moments of the prior distributions for the parameters $A_1, ..., A_p, C$ and the residual covariance matrix Σ can be computed as

$$E[(A_1, ..., A_p, C)] = \bar{A} = (X^{D'} X^D)^{-1} X^{D'} Y^D, \tag{20}$$

$$Var[(A_1, ..., A_p, C)] = \bar{V} = (X^{D'} X^D)^{-1}, \tag{21}$$

$$\bar{S} = (Y^D - X^D \bar{A})'(Y^D - X^D \bar{A}) \tag{22}$$

and

$$\bar{\alpha} = T^D - n(p-1) - 1, \tag{23}$$

where T^D is the number of rows of Y^D.

Conditional on the most recent Gibbs draw i of the state vector Z_t^{i-1}, $A_1^i, ..., A_p^i, C^i$ and Σ^i are sampled from their respective posterior distributions. In particular, the $(A_1, ..., A_p, C)$ follow a multivariate t-distribution with mean $\tilde{A}$, covariance matrix $\tilde{S} \otimes \tilde{V}$ and degrees of freedom $\tilde{\alpha}$, while $\Sigma \sim IW(\tilde{S}, \tilde{\alpha})$. The respective moments of these distributions can also be computed using the dummy variables outlined in equation (19). In particular, augment the state vector Z_t^{i-1} with the dummy variables to obtain $Y^* = [(Z_t^{i-1})', Y^{D'}]'$ and $X^* = [(Z_{t-1}^{i-1})', X^{D'}]'$. Then

$$\tilde{A} = (X^* X^*)^{-1} X^* Y^*, \tag{24}$$

$$\tilde{V} = (X^* X^*)^{-1}, \tag{25}$$

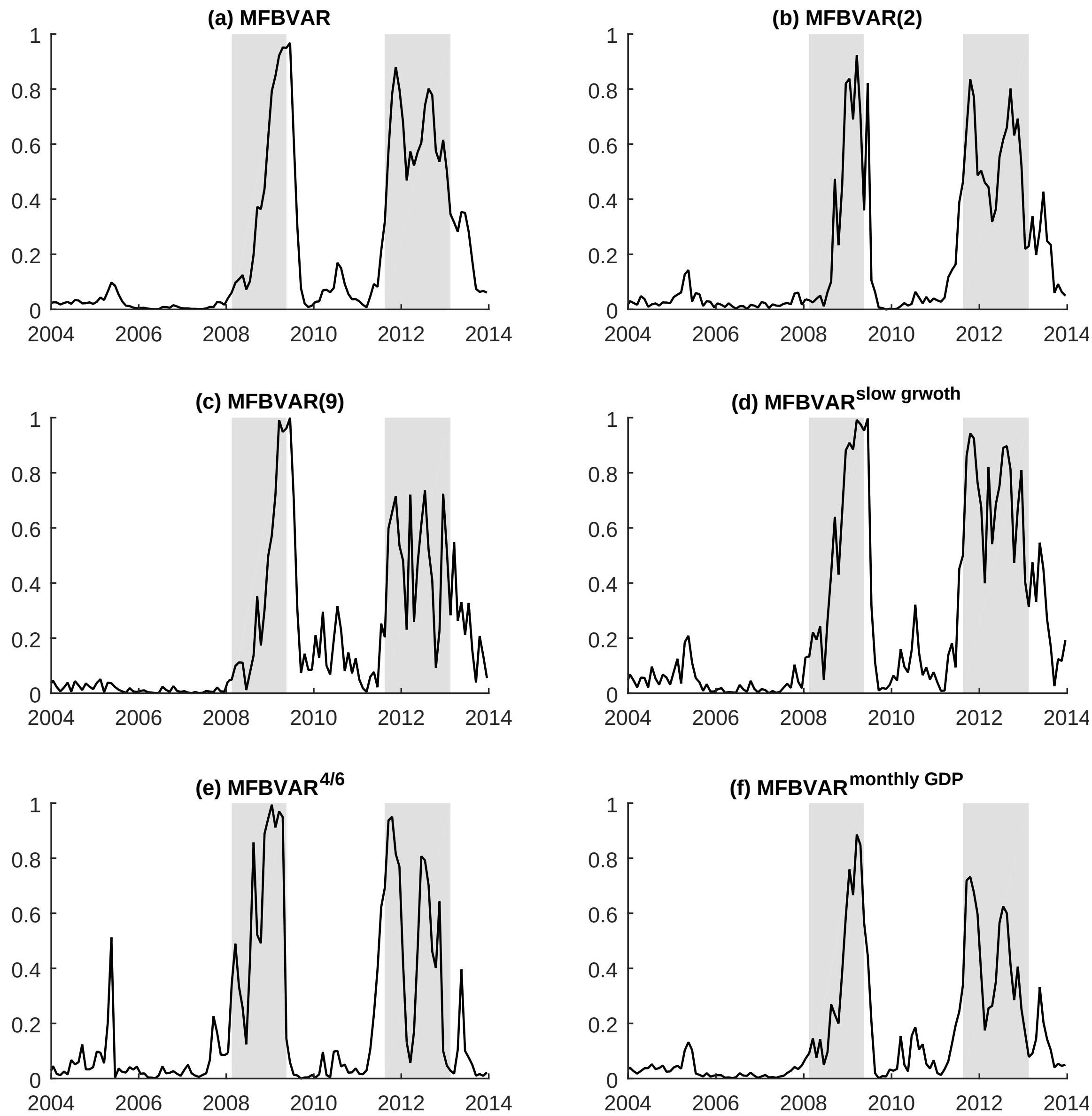
$$\tilde{S} = (Y^* - X^* \tilde{A})'(Y^* - X^* \tilde{A}) \tag{26}$$

and

$$\tilde{\alpha} = T^* - np + 1, \tag{27}$$

where T^* is the number of rows of Y^*.

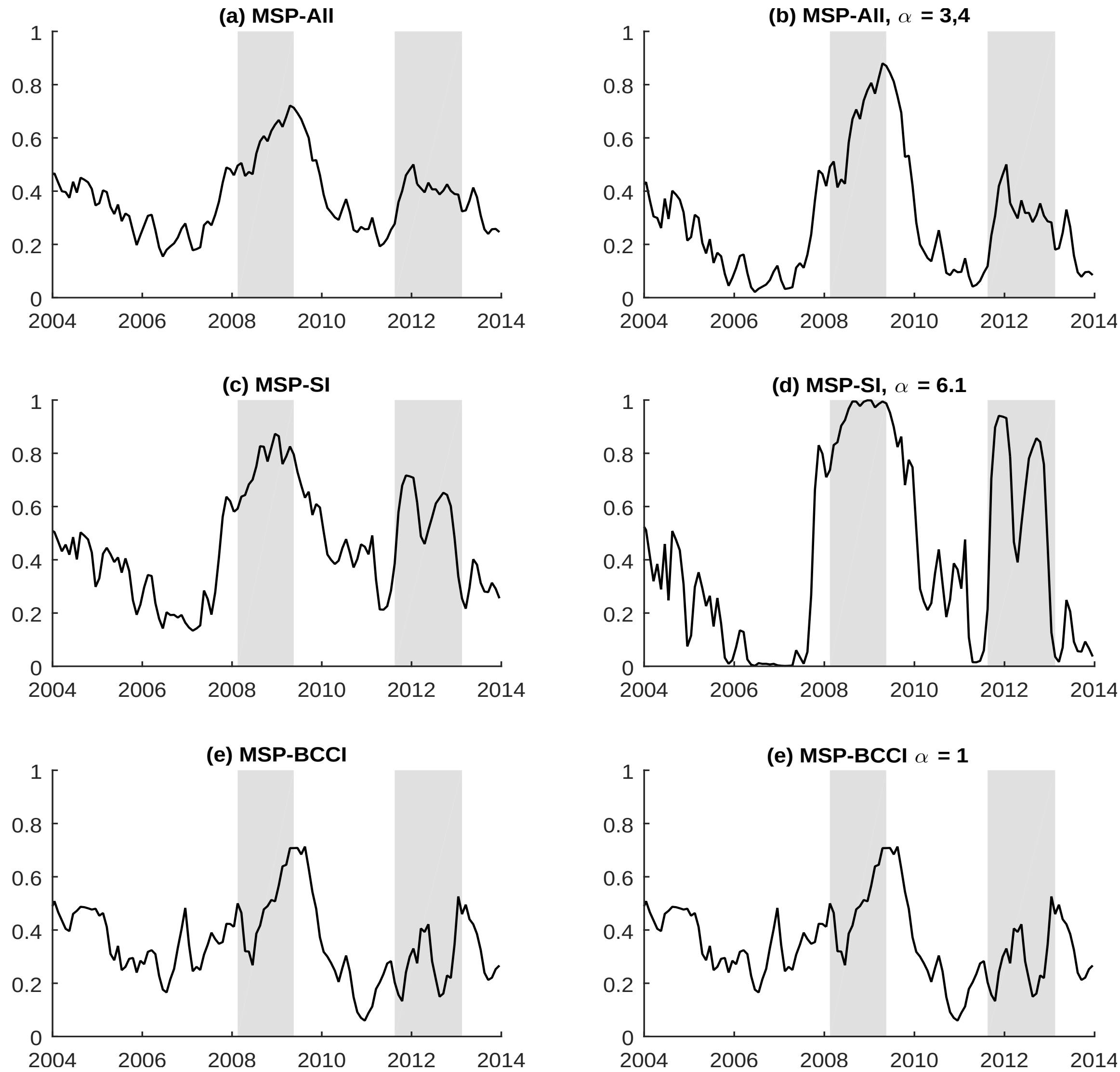
A 3: Alternative MFBVAR Real-Time Recession Probabilities



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas denote euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. For the model abbreviations see section (7).

Figure 5: Real-time recession signals for the euro area, MFBVAR robustness.

A 4: The Beta Transformed Linear Opinion Pools



Notes: The monthly real-time recession signals displayed are computed as three-month weighted moving averages over the original probabilities obtained with the alternative approaches (see footnote 10). The shaded areas denote euro area recessions as dated by the CEPR Euro Area Business Cycle Dating Committee. For the model abbreviations see section (7).

Figure 6: Real-time recession signals for the euro area, Pool robustness.