Anticipating business-cycle turning points in real time using density forecasts from a VAR

Sven Schreiber[a,b,*], Natalia Soldatenkova[c]

[a] *Macroeconomic Policy Institute (IMK), Hans-Böckler-Str. 39, Düsseldorf 40476, Germany*
[b] *Free University of Berlin (Institute of Statistics and Econometrics)*
[c] *University of Bonn (Bonn Graduate School of Economics, BGSE), Lennéstraße 35, Bonn 53113, Germany*

Abstract

For the timely detection of business-cycle turning points we suggest to use medium-sized linear systems (subset VARs with automated zero restrictions) to forecast monthly industrial production index publications one to several steps ahead, and to derive the probability of the turning point from the bootstrapped forecast density as the probability mass below (or above) a suitable threshold value. We show how this approach can be used in real time in the presence of data publication lags and how it can capture the part of the data revision process that is systematic. Out-of-sample evaluation exercises show that the method is competitive especially in the case of the US, while turning-point forecasts are in general more difficult in Germany.

Keywords: density forecasts, business-cycle turning points, real-time data, nowcasting, bootstrap

JEL codes: C53 (forecasting models), E37 (cycle forecasting)

---

1. Introduction

In this paper we suggest a linear system approach to the old problem of business-cycle turning-point prediction, taking into account the data availability and revision problems in real time[1]. This approach differs from the usual methods used to detect business-cycle turning points, which is usually done with non-linear models such as probit or Markov-switching methods.[2] The general idea is that we use a linear (system) model to predict the continuous output variable several steps

*Corresponding author
Email address: `svetosch@gmx.net` (Sven Schreiber)

[1] The subVAR applications were carried out with the *gretl* econometrics software (see Cottrell and Lucchetti, 2013, especially chapter 8 on real-time data), the model confidence set procedure was done with the MulCom package for Ox, and the comparison models were implemented in Matlab

[2] While a subclass of Markov-switching models can actually be regarded as linear, the general case –for example with regime-dependent dynamics– yields non-linear models (see Krolzig, 1997).

ahead. We then use the estimated probability density function (pdf) of the forecast to calculate the probability of a realization below the previously defined recession threshold (or above a certain boom threshold). As described in more detail below, using monthly data we employ a threshold of a negative cumulated growth rate of -1% over a time span of six months to call a recession.[3]

Our approach has the following advantages: First, we can define the real-time variables in our multivariate system such that we also capture the revision process of consecutive data publications, by keeping (some of) the superseded data publications in the econometric system.[4] Secondly, compared to probit models there is reason to hope that the direct forecast of the continuous output variable is better able to exploit the information contained in the data. After all, in order to fit the probit model it is necessary to reduce the target variable to a binary regime variable, which discards quite a bit of information. Because of the linearity of the estimator our method is also computationally robust. Finally, as shown in the applications below, considering more than two regimes just means to partition the predictive density into more than two exogenously defined regions, which is straightforward.

There are also drawbacks of our approach which have to be acknowledged: In order to make use of a broad information set, we use relatively large VARs as the starting point for our forecasting model. These initial models are then reduced with automated coefficient restrictions following the general-to-specific method, but the initial models suffer from the curse of dimensionality, i.e. the combination of too many variables and lags may exceed the available degrees of freedom. In a scenario with only quarterly data and only short available revision data histories our approach may therefore not be the most suitable one. A more fundamental restriction is that our model presupposes linear time series processes. Thus if the DGP were actually non-linear, our forecasting models would only be approximations. On the other hand, the same variables are often analyzed with linear models in other macroeconomic contexts, and thus linear models seem to be perfectly reasonable.

Finally, our method is also affected by a turning-point recognition lag. If for example we receive in some publication period a recession signal based on a forecast h steps ahead (realistically assuming a moderate forecast horizon $h < 5 + p$, where p is the publication lag, i.e. the number

[3]The averaging of consecutive months can be naturally interpreted according to the "triangle" approach by Harding and Pagan (2002), where episodes can be very short and intense or more drawn out and gradual to qualify as recessions.

[4]See Corradi, Fernandez, and Swanson (2009) for a discussion of the information content of data revisions.

of periods it takes before an initial data release happens), this means that the beginning of the recession actually happened in some reference period up to $-(h - p - 5)$ months ago, and so in reality the recession would likely be already underway. Although this may seem unfortunate, it is the logical consequence of the definition that a decline in economic activity must have a certain minimum duration to call it a recession.[5]

Related literature. An early example that linear prediction models can be applied to the problem of turning point determination with continuous target variables is given in Stock and Watson (1993). Österholm (2012) uses a similar approach as we do, in the sense of applying a linear model (Bayesian VAR, BVAR), and working with the predictive densities. We discuss and apply his approach in section 2.

Non-linear turning point applications are an active area of research; in the domain of models with binary dependent variables the introduction of an explicitly dynamic probit setup by Kauppi and Saikkonen (2008) has spurred applied research, for some recent examples see Ng (2012), Nyberg (2010), and Proaño and Theobald (2014). The main alternative is Hamilton's Markov-switching approach, and some recent applications with real-time data are Hamilton (2011), Nalewaik (2012), or Theobald (2014).

Complete real-time datasets with various data vintages are not as readily available as standard final-release data. Therefore we start our study in the next section with a slightly longer pseudo real-time dataset of the USA for 1986-2013 where we take into account the data publication lags, but we do not include the revision information. With this simplified dataset we introduce our suggested approach and by means of out-of-sample forecasting we compare it with other established methods for the determination of turning points, namely a Markov-switching model, several dynamic probit specifications, and the BVAR mentioned above. After that section the principal workings of our method should be sufficiently clear and we proceed in section 3 with the analysis of actual real-time data sets with revision information. The available samples in that case are 1993-2013 for the USA and 1995-2012 for Germany which are used for fully real-time applications of our approach in section 4.

[5] In probit or Markov-switching models a formal minimum-duration requirement is typically missing for the forecasts. Instead at estimation time a specification is chosen that somehow delivers reasonable regime classifications in sample, where "reasonable" usually also means that the regime episodes should not be too short.

2. Pseudo real-time simulation: method and application to the USA

In section 3 with the various actually published vintages of the data we explain more details of our proposed subset VAR (subVAR) approach. In contrast, in the present section we compare a simplified version of our approach with other methods that have been used in the literature to detect turning points. The simplification consists of using only final data releases (as of 2015) for all historical variables, not the vintages that were actually available in real time. However, the various econometric models still take into account the publication lags of the variables: for example, the posited information set for period t never includes the realization of industrial production in t, because a first release of that observation would only be available from $t+1$ onwards. Our simplification in this section means that instead of using the first vintage published in $t+1$ we use the revised value published much later (but still describing activity in period t, of course). We call this exercise a pseudo real-time evaluation, which for reasons of easy data availability we conduct in the sample 1986m2 through 2013m4, where the initial estimation sample ends in 1999m12. Throughout this section we work with variables that are either stationary or have been transformed to stationarity by forming (log) differences. See the section 3 for a discussion of how to use non-stationary levels in our proposed approach.

Apart from the industrial production index, the dataset we use includes new orders of durable goods, the CPI, oil prices, a stock index, real loans (volume), the conference board composite index, a corporate bond yield spread, and the 3-months T-bill rate, which are all non-stationary and are transformed to growth rates or changes. As stationary variables we have the Michigan confidence index, the number of building permits, and a term spread.

The other approaches that we consider are established in the literature, namely a Markov-switching single-equation model, another single-equation model with a dynamic probit specification, and another VAR model where the shrinkage of the parameter space is achieved by using a Bayesian approach instead of our subset restrictions in a frequentist tradition. We now explain the most important features of these other models.

2.1. Established approaches

The univariate Markov switching (MS) model follows the exposition in Hamilton (2011) and Chauvet and Hamilton (2006), but now using monthly industrial production growth instead of quarterly GDP growth; see online appendix A.1 for more information. Hamilton and co-authors effectively perform a backcast, because they used the data vintages that are released four months after the respective quarter (q), thus they recommend working with data more recent than the following quarter ($q+1$). We follow them and also produce a backcast with the MS framework in our monthly setting. That is, with the data released at the current edge T (which in the case of industrial production refers to period $T-1$) we estimate the probability of a recession in $T-1$.

The form of the dynamic probit (DP) model used here was suggested by Kauppi and Saikkonen (2008). The binary target variable in our case is the NBER recession indicator, see online appendix A.2 for model details. For the inclusion of a lagged NBER indicator term in the equation the publication lag of the NBER business-cycle dating committee plays a special role, as the committee make their announcements at non-fixed intervals and often quite long after the fact; the delay of announcing the start or end of recessions has ranged from 5 to 20 months. For our specification of the probit equation we make the extreme assumption that the NBER indicator values are always published already after five months. Imposing the historical minimum delay is obviously counterfactual and would be impossible in real time, but treats the probit specification quite favorably.[6] In terms of explanatory variables we consider two specifications: in the first one only the available data on industrial production growth is included, and in the second equation all variables that also appear in the VARs below are used. Lag orders are chosen through the BIC. With these probit models we produce recession probability forecasts ranging from a backcast (e.g. data releases from T, which refer to $T-5$ in case of the NBER indicator, and to $T-1$ for industrial production, are used to estimate the probability in $T-1$) through a forecast four months ahead (data released in T are used to estimate the probability in $T+4$).

Closest in spirit to our own approach is the use of a Bayesian VAR (BVAR) to derive and evaluate a forecast density. The general setup we use here stems from Villani (2009), and Österholm (2012) applied it to the turning-points forecasting problem (cf. online appendix A.3), but he found that it performs quite badly in a quarterly GDP-based context. We follow them in setting the lag decay parameter of the prior to unity, the overall tightness to 0.2, and the cross-equation tightness to 0.5. Österholm defined a recession as two consecutive quarters with negative growth, but in our monthly context we apply the same definition and threshold as in our subset VAR, see below. Since it is possible that the small dimensionality of his system was responsible for the disappointing results we consider the bigger information set that we also use for our subset VAR approach; see also Banbura, Giannone, and Reichlin (2010) for the merits of bigger BVARs for forecasting. The forecasting steps produced with this model range from a

[6] An alternative solution would be to construct another binary recession indicator in real time, see Proaño and Theobald (2014). In any case, the dynamic probit model would gain considerably from a shorter availability lag of the target indicator, but that problem is unavoidable in reality.

nowcast (using data released in T to determine the probability in T) to four months ahead.

2.2. The simplified subset VAR approach

In this simplified approach we do not include the revision information or various data vintages, which allows us to keep the notation simple. The model is in principle a standard VAR, $y_t = A(L)y_{t-1} + Md_t + u_t$, where the time index of the variables denotes the dates of initial data publication, and d_t is a vector of a constant term and seasonal dummies. Thus the n-dimensional vector y_t consists of the values that are published in period t, and to which periods these values refer depends on the different publication lags of the components of y_t (e.g. for industrial production it refers to $t-1$). The knowledge of the publication lags is a piece of meta information here.

The fundamental idea is to forecast the entire sequence of probability density functions (pdf) of industrial production $h = 1...H$ steps ahead, where the steps refer to the econometric model in terms of the data publication time. For example the VAR forecast steps 1 to 5 refer to actual real-time forecast steps f from 0 (nowcast) to 4 months ahead. From the joint density of all forecasting steps the density of the cumulated growth rates spanning 6 months (two "rolling" quarters) can be derived, e.g. for a forecast step h we would cumulate the observed growth rates published in $T-(6-h)+1$ through T and add to them the forecast growth rates from $T+1$ to $T+h$. Because the forecast pdf at various horizons are correlated it is natural to choose a bootstrap approach to simulate the relevant densities. Specifically we resample from the estimated model residuals to simulate the innovations, such that we also allow for non-Gaussian innovations u_t.

In principle the model can be re-estimated every time the sample is expanded. To save on computation time, however, we re-estimate the parameters only every three observations and simply update the residual vector in between.

2.2.1. Subset restrictions

From the impulse response function (IRF) literature we know that allowing for longer lags can be useful. The intuition given by Kilian (2001) is that higher-order polynomials may capture the multi-step curved dynamics better. Obviously, the problem of multi-step forecasts has some closely related aspects to the IRF problem, and therefore we follow Kilian in favoring the AIC information criterion over the stricter BIC criterion. However, in contrast to typical structural VARs our systems include many more variables, and thus the long lags introduce serious estimation inefficiencies, which requires some sort of shrinkage method. The BVAR approach mentioned before would be one solution, and indeed it turns out that with the recession definition explained in section 2.2.3 and in this monthly context the BVAR fares better than in Österholm (2012). Nevertheless, here we propose a subset restriction search instead which turns out to be quite successful.

Our estimation procedure for each sample is initialized by OLS, then the redundant regressors are removed one by one based on their t-ratios.[7] This procedure is repeated until no regressor has a p-value above the selected threshold α_{cut}, where we have set $\alpha_{cut} = 0.01$. Retaining more coefficients with a more liberal significance level tended to deteriorate the results, given that there still remains a considerable number of coefficients to be estimated. At the end the entire system is re-estimated efficiently with the SUR method.

2.2.2. Innovation and parameter uncertainty

The unsystematic out-of-sample innovations u_t, $t = T + 1...T + h$, are the main source of forecast uncertainty. However, another source of forecast uncertainty in finite samples is given by the fact that the model parameters are also subject to sampling uncertainty. We have addressed this concern by implementing another bootstrap layer to simulate the joint distribution of the model parameters. The algorithm works as follows, given a pre-specified maximum lag order of p_{max}:

1. For a sample with final data release in T, denoting the available data matrix with Y_1 ($T + p_{max}$ by n), determine the lag order $p \in \{0, ..., p_{max}\}$ of the unrestricted VAR by the AIC criterion.
2. Estimate the parameters of the unrestricted VAR(p) with the data Y_1. Denote the estimates with β_1, and the corresponding residual matrix (T by n) with U_1.
3. Resample from U_1 to obtain a new innovation version U_2.
4. Simulate a new version of the data Y_2 by using the parameters β_1 and the simulated innovations U_2.[8]
5. Perform the subset restriction search on the VAR(p) applied to the new data Y_2 as described in section 2.2.1, using a prespecified cutoff significance level α_{cut}, and denote the resulting parameter estimates by β_2. The corresponding residuals from the restricted system are called U_3.

[7] For computational efficiency we perform this elimination separately for each equation. It would be straightforward to conduct it on the system level which in practice is not necessarily superior, however. For a comprehensive discussion see Brüggemann (2004).

[8] There are various possibilities how to choose initial values. Here we also draw them randomly.

6. Using the final in-sample observations of the original data Y_1 from $T-p+1$ through T as initial values for the forecast, simulate the system forward H steps with the parameters β_2, and using new innovations U_H (H by n) that were obtained by resampling from U_3. Denote (and store) the resulting simulated realizations F (H by n).
7. Calculate the cumulated 6-month growth rates for the simulated realizations of industrial production from $T-(6-h)+1$ to $T+h$, for $h=1...H$, using the original data Y_1 up to T and the simulated values F after T.
8. Repeat steps 3 through 7 many times to approximate the densities of all forecasts up to horizon H.

At the core this method is a straightforward residual-based bootstrap appropriate for such time series, see Fresoli, Ruiz, and Pascual (2015) for a recent formalization. Notice that step 5 amounts to something very similar to the endogenous lag order bootstrap proposed by Kilian (1998), because in each bootstrap iteration the subset restrictions on the parameter vector will in general be different, including the possibility that coefficients of higher-order lags could all be set to zero.

In steps 3 and 6 one could also employ a (residual-based moving) block bootstrap to account for any remaining serial correlation of the residuals. We have also checked such a variant but the results were not noticeably different, wherefore we only report the results from the standard *iid* bootstraps. For the lag order of the unrestricted system we use the same value $p=6$ that is chosen by the AIC in the full real-time application in section 3 below.[9]

2.2.3. Determination of the recession threshold

A central parameter of the subVAR approach is a threshold value to partition the forecast density in a recessionary and a non-recessionary part. Figure 1 (upper panel) illustrates the behavior of the US industrial production 6-months growth rate in relation to the NBER recession dates for each reference period of the sample 1986m01-2013m12 using final-vintage data. The 6-months growth rate of industrial production matches quite well the NBER recession dates considering a threshold of -1%.[10] It is also apparent that in the industrial sector episodes of

[9] In this longer sample it would be possible in principle to pre-specify a higher maximum lag order p_{max} of 7 or 8, and indeed in this sample the AIC then always recommends this maximum value. However, the computational burden of this bootstrap-within-bootstrap then becomes an issue for such a long evaluation sample because the number of coefficients grows exponentially with the lag length. Therefore we did not increase the lag length beyond 6. Of course the computational cost is much less of an issue (by a factor of at least 50) if a single set of forecasts is produced in actual real time at the current edge instead of an out-of-sample evaluation.

[10] Exceptions are the episodes in 1989m7, 1989m10 and 2005m9. In 1986m6 the realized growth rate comes close to the threshold (-0.00952).

temporary contractions with 6-months growth rates slightly below zero are quite common, and thus it makes sense to use a negative threshold to call a recession.

Alternatively one could use additional indicator variables to specify the recession definition. In the lower panel of figure 1 the 6-month change in the US unemployment rate is shown. The idea is to combine information from the industrial production and the unemployment rate variables into one recession criterion such as:

$$\alpha\, g(IP) - (1-\alpha)\, g(UR) < thr, \tag{1}$$

where $g(IP)$ and $g(UR)$ are the 6-month growth rates or changes and thr is the recession threshold. We calibrate the values of α and thr parameters for the full sample, where α is allowed to change from 0 to 1 with 0.01 step while thr changes from 0 to -2% with step -0.1%. The results of the calibration show that without the unemployment rate (i.e. with $\alpha = 1$), the optimal threshold for identifying both recession and expansion periods is -1% with 94.7% of correct classification. Adding the unemployment variable to the calibration selects $\alpha = 0.2$, $thr = -0.8\%$, and yields 96.2% classification success.[11] These results show that adding the unemployment rate to the recession definition provides only marginal improvements in terms of identifying recessions and expansions. Therefore we choose to retain the simple recession definition based on the 6-month growth rate of industrial production alone, with a relevant threshold of -1%.

2.3. Results (pseudo real time)

In figures 2 to 5 the various estimated recession probabilities for the pseudo real-time simulation are shown. When comparing the fit to the NBER benchmark across models it should be borne in mind that a lower-horizon forecast is generally easier, and below we discuss how this affects the timing of the signal publication.

It is noteworthy that the estimated probabilities of the Bayesian VAR (figure 4) appear compressed towards the center of the unit interval, never falling below 20% and never rising over 95%. However, an interesting finding is that the BVAR model is considerably more successful here than it appeared in Österholm (2012), where he tried to forecast turning points of GDP growth and found extremely low recession probabilities throughout. We could qualitatively replicate his

[11]The same calibration can also be conducted for the estimation sample up to 1999m12 which we later on use for in-sample model fitting. The results with both variables are $\alpha = 0.1$, $thr = -0.5\%$ and 97.6% of cycle phases are identified correctly, while for the specification with the IP alone ($\alpha = 1$) the recession threshold is again -1% and 95.8% of turning points are recognized.

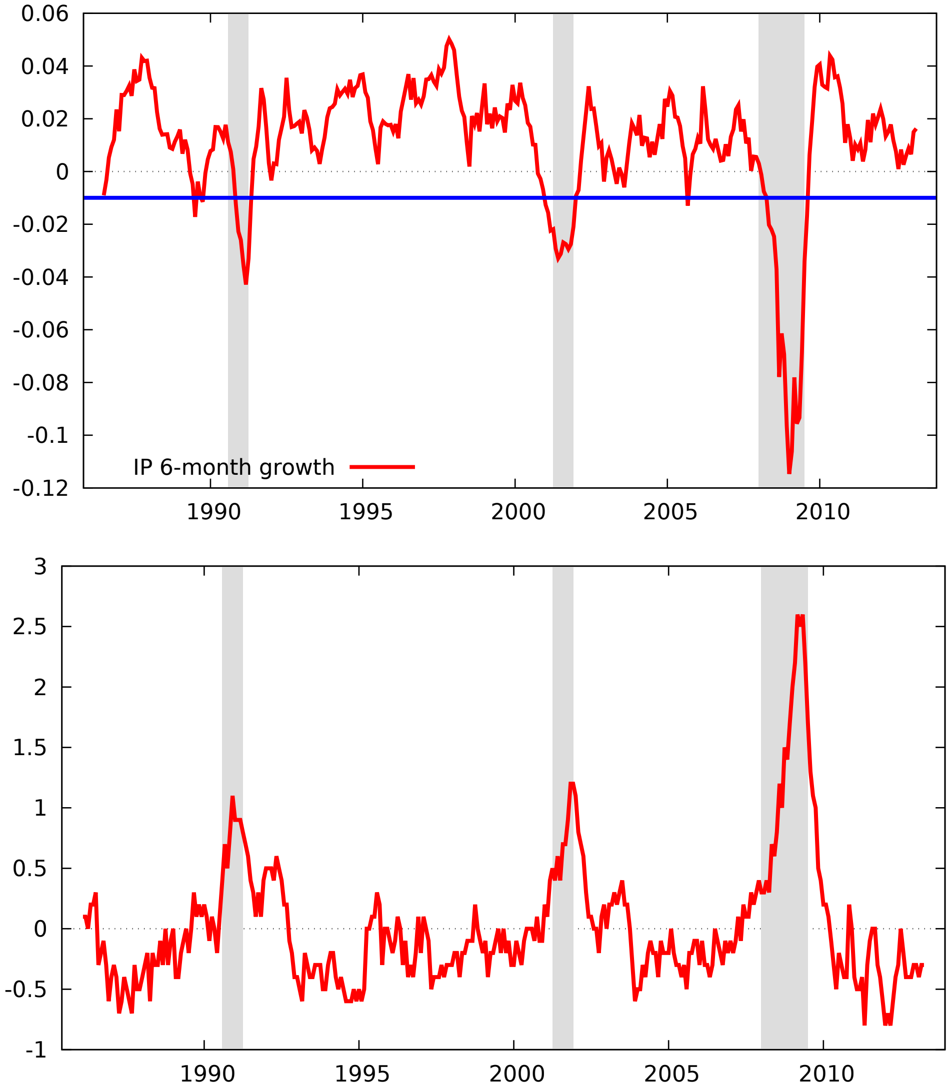

Figure 1: US industrial production, 6-month growth rate (upper panel), and US unemployment rate, 6-month change (lower panel). Sample period 1986m2-2013m4. The time axis corresponds to the reference period. The straight blue line is at -1%. The shaded regions correspond to NBER recession dates.

findings when we used a recession definition of six consecutive months with negative IP growth in each of them. In contrast, our definition relies on very low cumulated growth which encompasses episodes where intermediate months may have slightly positive growth.[12]

The earlier recession in 2001 would have generally been difficult to recognize in a timely fashion. The MS backcast model is a special case in that it tended to see the recession occurring earlier than the NBER later decided, while other models shift the start and end to the right, including the Probit backcast models. Exceptions are the VAR-based nowcasts which match the NBER timing quite well (subVAR $f = 0$, and also BVAR $f = 0$ if we abstract from the general "compression" of the probability range in the BVAR models). The recession probabilities from the higher-horizon forecasts (Probit $f = 4$, BVAR $f = 4$, subVAR $f = 4$) never actually rise above 50% in 2001. These higher-order forecasts are also problematic in the great recession in 2008/2009 after the financial crisis, thus we can confirm that attempting turning point forecasts more than one or two months ahead is ambitious.

Around the end of 2005 many models display recession probabilities above 50%. According to the NBER dating this would have been a false positive, but indeed we could already observe in the ex-post data in figure 1 that industrial production 6-month growth was below -1% then.

The fit of the various curves in figures 2 to 5 is evaluated with standard forecast accuracy measures in table 1. Once again it turns out that it is quite hard to beat the benchmark of an AR(1) model in growth rates, which delivers an effective nowcast ($f = 0$) by performing a 1-step forecast on the IP data with a publication lag of one period. For the AR(1) benchmark the recession definition is the same as for the VARs. The Markov-switching approach fully uses its advantage as a backcasting model and achieves a reduction of the error measures below 80% of the AR(1) values. The only non-backcasting model that is capable of improving upon the AR(1) benchmark is the nowcast of the subVAR. The subVAR also has the best fit among the models with 2- and 4-months ahead forecasts, except for the MAE with $f = 4$ where it comes in second behind the Probit.

Table 2 reports a summary of the timing of the signals coming from the various models. As a benchmark model at the top we report again the pure AR(1). Note that while in this table we use 50% as a natural threshold for the probabilities to call the start or end of a recession, other thresholds have also been used in the literature (cf. Hamilton, 2011, or Ng, 2012). It is clear

[12]This is not a criticism of Österholm (2012) because in his quarterly context the definition of two quarters with negative growth was natural.

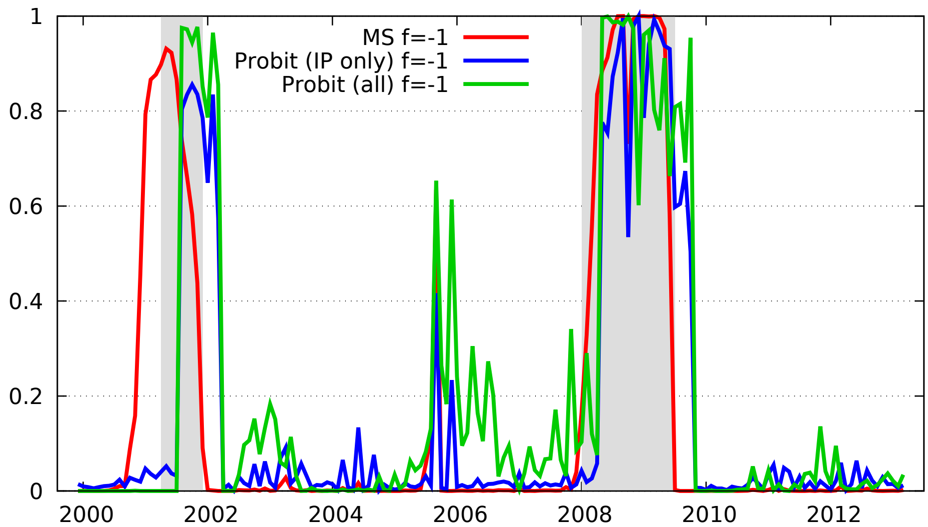

Figure 2: Markov switching (MS) and dynamic probit models, estimated probabilities of a recession at effective forecast step $f = -1$ (backcast). The first probit model contains only the NBER indicator (with an assumed publication lag of 5) and industrial production, the second probit model includes all predictors. The time axis denotes reference periods, shaded areas are NBER dates peak through trough.

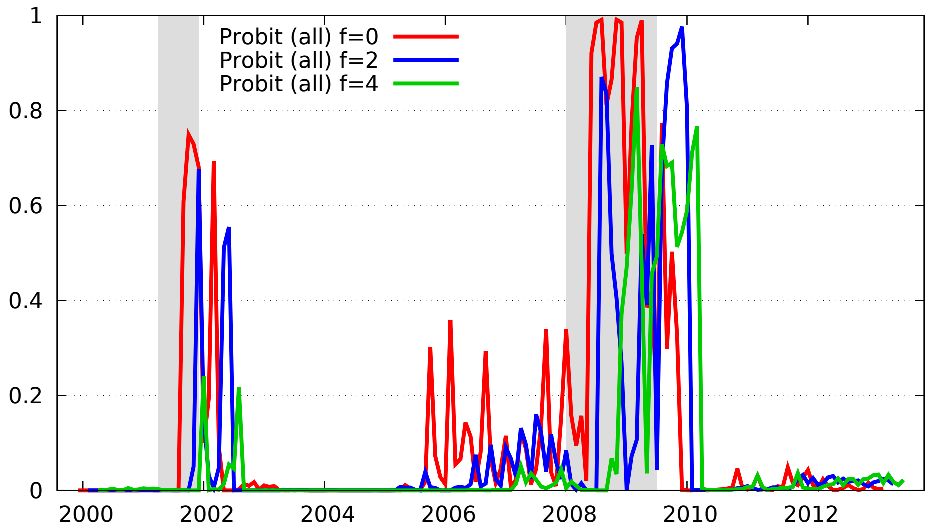

Figure 3: Probit models (dynamic), estimated probabilities of a recession at effective forecast steps $f = 0$ (nowcast) and $f = 2, 4$. The models contain all predictors. The time axis denotes reference periods, shaded areas are NBER dates peak through trough.

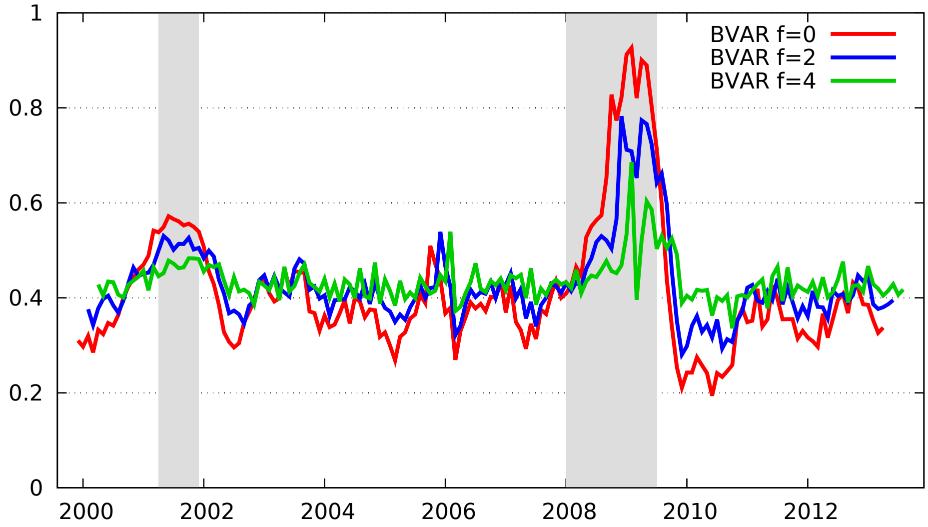

Figure 4: Bayesian VAR models, estimated probabilities of a recession at effective forecast steps $f = 0$ (nowcast) and $f = 2, 4$. The models contain all predictors. The time axis denotes reference periods, shaded areas are NBER dates peak through trough.

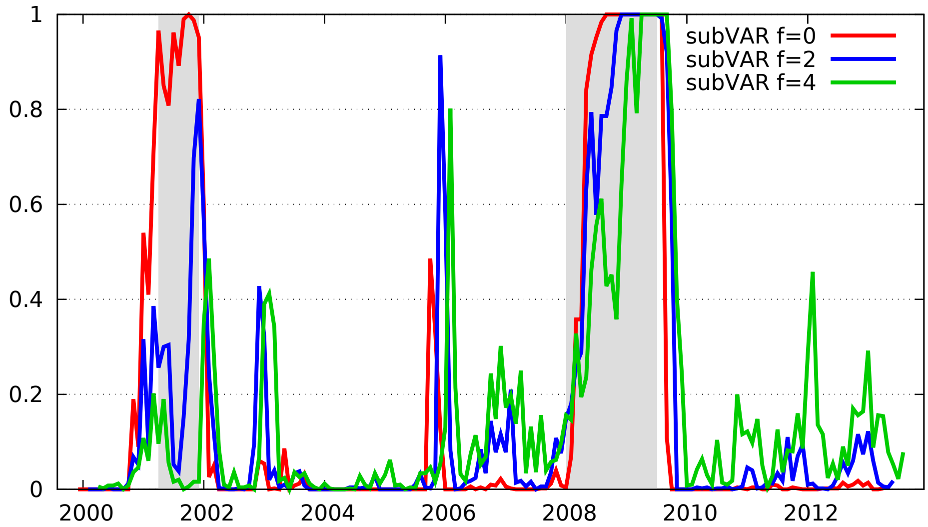

Figure 5: Subset VAR models, estimated probabilities of a recession at effective forecast steps $f = 0$ (nowcast) and $f = 2, 4$. The models contain all predictors. The time axis denotes reference periods, shaded areas are NBER dates peak through trough.

Table 1: Evaluation measures of the pseudo real-time simulation

	effective forecast step f	Root mean squared error (RMSE)	mean abs. error (MAE)	TheilU1
AR(1) (growth rate)	0 (nowcast)	0.239	*0.071*	0.293
Markov-Switching	−1 (backcast)	*0.76*	*0.85*	*0.80*
Probit IP only	−1	1.15	1.65	1.26
Probit all indicators	−1	1.28	2.04	1.33
	0	1.19	1.78	1.38
	2	1.67	2.71	2.12
	4	1.67	*2.67*	2.27
BVAR	0	1.59	5.20	1.53
	2	1.75	5.82	1.68
	4	1.86	6.25	1.80
subVAR	0	*0.91*	1.01	*0.92*
	2	*1.24*	*1.96*	*1.35*
	4	*1.45*	2.79	*1.68*

Notes: All results relative to the AR1 nowcast (at the top), but comparisons between different forecast horizons should take into account that lower-horizon forecasts (including backcasts) are inherently easier. The best value across models for a given step f is given in *italics*.

The concrete formula for this version of Theil's U is $TheilU1 = \sqrt{\overline{(y_t - \hat{y}_t)^2}} / \left(\sqrt{\overline{\hat{y}_t^2}} + \sqrt{\overline{y_t^2}} \right)$ which is defined even if some y_t values of zero, as is the case with the NBER recession indicator here.

from figure 4 that using a different probability threshold would especially affect the conclusions from the BVAR model, less so for the other models.

The ideal model would implement a large effective forecasting horizon f and would accurately anticipate events. In such a model the signal would be published exactly f periods before the fact, thus the signal shift s would be just minus the forecast horizon, $s + f = 0$. Given that such an ideal model is rarely available, it turns out that no single model performs best overall, and there is a trade-off.

On the one hand one could favor early signals ($|s|$ large, $s < 0$) at the cost of also expecting the turning point before or after it actually happens ($s + f < 0$ or > 0). For this preference a first example in table 2 is the Markov-Switching (MS) model for the start of the recession after the peak 2001m3: The recession signal would have occurred two months before the recession started, but given that the MS model is based on a backcast, the signal would have meant that the recession was already well underway when according to later NBER announcements it had not started yet ($s + f = -3$). Another example relates to the start of the great recession after the peak 2007m12: the subset VAR (subVAR) would have provided the earliest published signal ($s = +2M$), but only in the specifications with effective forecast steps $f = 2$ and $f = 4$, thus expecting the turning point much later than in reality ($s + f = 4$ or $s + f = 6$). The flip side of the trade-off is to accept a long turning-point recognition lag (large s) while aiming for an accurate turning-point date ($s + f = 0$). The leading example of such a preference is the NBER methodology itself which is an extreme backcast.

All in all, it is remarkable how well simple model specifications work which rely on the industrial production as the only predictor. This comprises the AR(1) benchmark (in growth rates), the univariate Markov switching model, and to a lesser extent the first dynamic Probit specification. Our proposed subVAR model is able to improve upon those results in some dimensions and especially at shorter forecasting horizons, but the gain is somewhat limited. It remains to be seen whether the inclusion of the revision process in real time helps to improve the subVAR further.

Notice that theoretically the dynamics of the revision process should indeed be relevant in the growth-rate specification, because due to the differencing of the original data we expect a moving-average type autocorrelation of the revisions, which would be implicitly modelled with lagged terms.

Table 2: Signals in the pseudo real-time simulation

	effective step f	Signal shift s at peak 2001m3	$s+f$	s at trough 2001m11	$s+f$	s at peak 2007m12	$s+f$	s at trough 2009m6	$s+f$	Further misclassifications
AR(1)	0	$-1M$	−1	$+2M$	2	$+4M$	4	$+2M$	2	2005m10
Markov-Switching	−1	$-2M$	−3	$\pm 0M$	−1	$+3M$	2	$+1M$	0	2005m10
Probit IP only	−1	$+5M$	4	$+5M$	4	$+5M$	4	$+5M$	4	–
Probit all predictors	−1	$+5M$	4	$+5M$	4	$+5M$	4	$+5M$	4	2005m10, 2006m1
	0	$+5M$	5	$+1M$	1	$+5M$	5	$-2M$	−2	2002m3, 2009m1, 2009m8, 2009m10
	2	$+6M$	8	$-1M$	1	$+5M$	7	$+5M^a$	7	2002m3, 2002m4, 2008m8 – 2009m1
	4	<no signal>	n.a.	n.a.	n.a.	$+9M$	13	$+5M$	9	2008m12 – 2009m3
BVAR	0	$-1M$	−1	$+2M$	2	$+4M$	4	$+2M$	2	2005m10
	2	$-1M$	1	$-1M$	1	$+4M$	6	$+1M$	3	2005m10
	4	<no signal>	n.a.	n.a.	n.a.	$+8M$	12	$+1M$	5	2005m10, 2008m11
subVAR	0	$-1M$	−1	$+2M$	2	$+4M$	4	$+2M$	2	2001m1
	2	$+5M$	7	$\pm 0M$	2	$+2M$	4	$+2M$	4	2005m10, 2005m11
	4	<no signal>	n.a.	n.a.	n.a.	$+2M$	6	$\pm 0M$	4	2005m10

Notes: Taking into account publication lags but abstracting from data revisions (final data vintages). Publication dates of estimates 1999m12-2013m4.

Here signals are defined as crossing 50% probability. A leading signal would be published up to (and including) the respective NBER peak/trough dates, and the signal shift s in months is negative for a signal date up to the peak/trough date, zero for a signal immediately after the peak/trough, positive for a lag. The effective forecast step f denotes the difference between the reference period about which the forecast is made and the real-time publication date of the information set.

The growth-rate AR(1) benchmark, the Bayesian VAR and the subset VAR all work with density forecasts and a recession definition of 6-month cumulated growth below −1%.

[a] Another interpretation would be a signal right before the trough with $s = -2$, but then all following months 2009m6– 2009m11 would have to be listed as misclassifications.

2.4. Further analysis of the subset VAR results

Next, we compare our results with the recent forecasts by Proaño and Theobald (2014, PT), which are derived from a different dynamic probit approach, and which they showed to be at least as good as other existing forecasts. Here we apply the Model Confidence Set (MCS) procedure suggested by Hansen, Lunde, and Nason (2011) which distills a set of models that are not significantly dominated by other models. The multiple tests are performed rigorously and do not suffer from pre-testing biases. If a MCS comprises more than one model, the different performances of the comprised models can be attributed to random sampling variability. Table 3 reports the various MCS results, where we consider two different target definitions and two different forecast error loss functions. The first target definition is the 6-month cumulated growth rate as before in this study, and the alternative definition is given by the NBER dating. The two functions squared and absolute loss that we apply to the forecast errors correspond to the RMSE and MAE criteria that were used above.[13]

The main result is that the forecasts from the subset VAR and from the PT dynamic probit models are often not significantly different from each other, or from the AR benchmark for that matter. Under squared error loss neither subVAR or probit dominate each other. For the case of absolute error loss the subset VAR has a slight advantage.

Finally, with our method we can also characterize the business-cycle outlook in more elaborate ways by differentiating regimes, because the problem of forecasting is separated from the problem of defining and specifying regimes, contrary to probit or Markov-switching methods. Here we split the non-recession region into a genuine “expansion” region and an intermediate “stagnation” region, fixing the additional threshold symmetrically at a six-month growth rate of +1%. We do not need to re-estimate anything, and the resulting three-region stacked probability plot is displayed in figure 6. This reveals for example that despite the low recession probabilities in late 2007 and early 2008, the chances for an expansion were also almost zero due to the very high stagnation probabilities.

[13]The fact the we use final data for this subVAR simulation may bias the competition somewhat in our favor. On the other hand, here we make no use of the revision information which will be seen to improve the subVAR performance further (section 3). A priori it is therefore not clear which model has an advantage here. The AR benchmark working with final data could have some advantage vis-a-vis the probit, however.

Table 3: Tests against existing other dynamic probit models

	target realizations by 6-month growth rate		target realizations by NBER dating	
effective forecast step	squared error loss	absolute error loss	squared error loss	absolute error loss
$f = 1$	dProbit, subVAR, AR (0.53)	subVAR, AR, dProbit (0.21)	dProbit, subVAR, AR (0.20)	subVAR
$f = 2$	subVAR, dProbit, AR (0.20)	subVAR	AR	AR, subVAR (0.24)
$f = 3$	AR, subVAR, dProbit (0.82)	AR, subVAR (0.95)	dProbit, subVAR, AR (0.13)	subVAR, dProbit, AR (0.13)

Notes: subVAR – subset VAR of this study, dProbit – dynamic Probit models from Proaño and Theobald (2014), AR – simple univariate autoregression. Each cell lists the models in the model confidence set (MCS), from best to worst. The evaluation sample is now 2000m1-2011m8 for comparison with the results from Proaño and Theobald (2014, their table 7). The analysis was performed with the MulCom package for Ox with a significance level of 10%. At the end the p-value of the respective worst non-rejected model in the set is reported in parentheses.

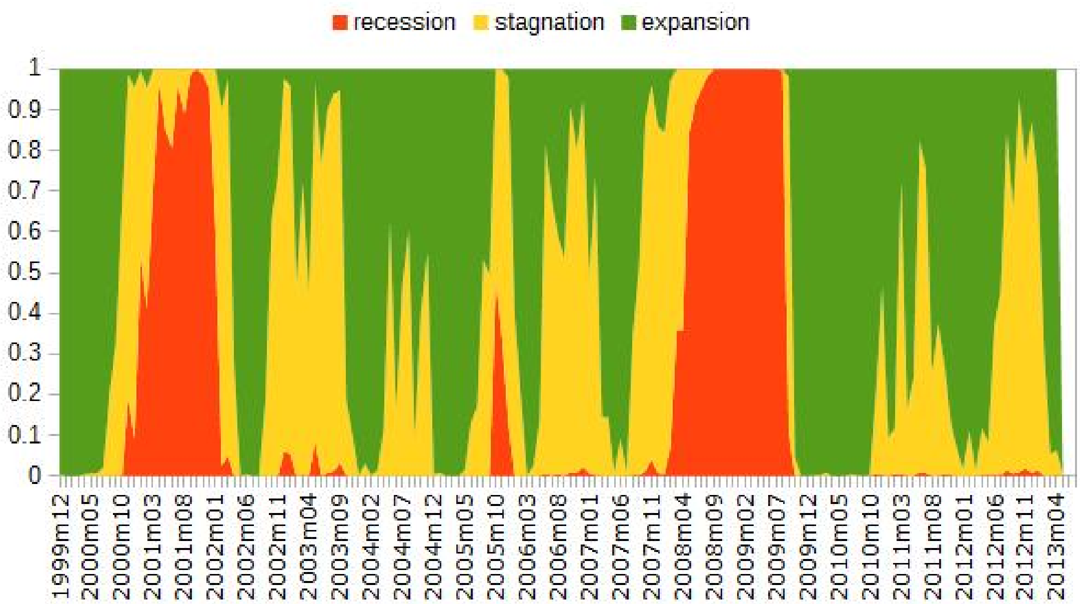

Figure 6: US three-regime probability plot, based on the subVAR nowcast.

3. Fully real-time forecasting

3.1. Notation

It is not so common to use VAR models for real-time data with publication lags and revisions, and thus we introduce some notation to capture these properties. We must distinguish two time axes: first the time period when a value is published, and secondly the time period when the measured activities actually took place. Throughout, we call the former "publication period" and the latter the "reference period". A well-known data presentation convention is by triangular matrices where the rows from top to bottom correspond to the reference periods (or "observations"), and the columns from left to right refer to the publication periods (or "vintages").

Initially we index every variable with both time axes, and thus every individual datapoint for use in the statistical model below is denoted by

$$x_t^{t-p-(r-1)}, \qquad r = 1, 2, ..., rmax;\ t = 1, 2, ..., T;\ p = const \tag{2}$$

where the publication period t is written as the subscript, and the reference period is put in the superscript. This reference period depends on the release number r and on the publication lag p, i.e. the number of periods it takes the statistical agencies to collect the input data and to publish their measurement. The publication lag can be different across variables and may be zero, but we assume that it is fixed over time for each variable.[14]

To give an example of the notation, if the publication lag up to the first release for a certain variable x is $p = 2$ months and we are talking about the second release $r = 2$, then the reference period is three months before the publication and the relevant datapoint is thus denoted with x_t^{t-3}. The absolute revision amount between the first ($r = 1$) and second ($r = 2$) release of such a variable with this publication lag $p = 2$ could be written as $x_t^{t-3} - x_{t-1}^{t-3}$.

The correspondence between our notation and the standard triangular-matrix representation is as follows, see also table 4:

[14]This is a potentially restrictive assumption, especially if we were using weekly or daily data, but for our monthly dataset and the relative recent samples it essentially holds. If there are isolated events when a first data release took longer than p periods to be published, the researcher could insert an artificial datapoint with a value that is extrapolated from past data. Also, we assume that there will be a data release in every month after the publication lag; if in reality there are some publication gaps, artificial pseudo-releases with unrevised values can be introduced in the dataset: $x_t^{t-1-p-(r_0-1)} = x_{t-1}^{t-1-p-(r_0-1)}$, where r_0 is the previous, actually available, release number.

- A certain column of the triangle, holding the data vintage published in period $t \leq T$: x_t^θ for $\theta = -p+1, -p+2, ..., t-p$
- A certain row of the triangle, holding the revision history for a reference period $\theta > -p$: x_t^θ for $t = \theta + p,\, \theta + p + 1, ..., T$
- The diagonal of the triangle, holding the respective initial data releases: x_t^{t-p} for $t = 1, 2, ..., T$

All of these items are data vectors where the meaning of "time" differs: in the first case the running index is for the reference periods, in the second case the running index relates to the publication time, and whenever diagonals are involved as in the third case, the time concept refers to both.

In real-time econometric studies, the researcher pays special attention to the available information set and thus to the data vintages at every period t. Due to possible data revisions an important difference with respect to standard time-series analysis arises whenever lagged values are involved. Typically, the lag operation still uses the current vintage from publication period t in the same matrix column, or formally $L_{ref}\, x_t^{t-p} = x_t^{t-1-p}$ for the first release $r = 1$ and with L_{ref} as the lag operator applying to the reference period axis. However, this lagged but vintage-t value will in general not be identical anymore to yesterday's information on the same reference period, x_{t-1}^{t-1-p}, because of the revision in the new release. Therefore it is relatively complicated to construct the relevant data matrices for use in econometric software packages.

In order to work with standard econometric methods and tools, we also consider the other diagonals of the triangular real-time data matrix representation. For each economic variable x we define the following collection $x_{1,t}^*, x_{2,t}^*, ..., x_{rmax,t}^*$ of statistical variables (where $rmax$ represents a cutoff point which is arbitrarily chosen, but in theory the revision process can go on forever):

1. The initial releases: $x_{r=1,t}^* = x_t^{t-p}$ for $t = 1, 2, ..., T$, which is identical to the main diagonal described above.
2. The second releases: $x_{r=2,t}^* = x_t^{t-p-1}$ also for all t, which yields the first sub-diagonal directly above the main diagonal.
3. ... and so forth until the $rmax$-th releases: $x_{r=rmax,t}^* = x_t^{t-p-(rmax-1)}$

The notable feature of this new representation is that each statistical variable now only has a single time index instead of two, and this index defines the available information set. This yields standard time series vectors and a standard lag operation can be unambiguously defined on the

Table 4: The real-time notation in relation to the triangular-matrix representation

publication periods → ↓ reference periods	2005m5, $t=1$	2005m6	2005m7	2005m8, $t=4$
2005m4	$x_1^0 = x_{1,1}^*$	$x_2^0 = x_{2,2}^*$	$x_3^0 = x_{3,3}^*$	$x_4^0 = x_{4,4}^*$
2005m5	na	$x_2^1 = x_{1,2}^*$	$x_3^1 = x_{2,3}^*$	$x_4^1 = x_{3,4}^*$
2005m6	na	na	$x_3^2 = x_{1,3}^*$	$x_4^2 = x_{2,4}^*$

Notes: The publication lag in this example is $p = 1$.

single time index: $L\,x_{r,t}^* = x_{r,t-1}^* = x_{t-1}^{t-1-p-(r-1)}$. It is this representation which enables us to use standard VAR tools.

A further advantage relative to the traditional real-time approach is that the information on the data revisions is kept in the model as part of the collection $x_{1,t}^*, x_{2,t}^*, ..., x_{rmax,t}^*$ (as long as some lags are included). For example, the revision $x_t^{t-3} - x_{t-1}^{t-3}$ is given by $x_{2,t}^* - x_{1,t-1}^*$ (in the case of a publication lag $p = 2$). Thus we have implicitly included a model of the revision process itself.

If a variable $x_{r,t}^*$ is stationary we include it in the $n_0 \times 1$ vector $\mathbf{s}_t$, and if it is integrated it belongs in the $n_1 \times 1$ vector $\mathbf{d}_t$,[15] where each revision-prone variable x will count as $rmax$ statistical variables in n_0 or n_1. The collection of the releases $r = 1, 2..., rmax$ of a certain variable x will either belong entirely into $\mathbf{s}_t$ or exclusively into $\mathbf{d}_t$. A corollary of this assumption is that the different releases of a certain economic variable which is integrated will automatically be co-integrated with unit coefficients.

The union of all variables is denoted in two separate ways:

$$\begin{aligned} \mathbf{z}_t^0 &= (\mathbf{s}_t^{'}, \Delta\mathbf{d}_t^{'})^{'} \\ \mathbf{z}_t^1 &= (\mathbf{s}_t^{'}, \mathbf{d}_t^{'})^{'} \end{aligned} \tag{3}$$

In $\mathbf{z}_t^0$ all variables are stationary, whereas the vector $\mathbf{z}_t^1$ is a mixture of stationary and integrated variables in levels. We refer to a model with $\mathbf{z}_t^1$ as the (log-) level specification, and to $\mathbf{z}_t^0$ as the growth rate specification. We are ultimately interested only in the outcome of log industrial production which we denote with y. Since (log) output is non-stationary, the collection $y_{1,t}^*, ..., y_{rmax,t}^*$ is a part of $\mathbf{d}_t$.

[15] The letter d stands for difference stationary. Other types of non-stationarity (trend stationarity, deterministic breaks) are not considered in our setup.

3.2. The subset VAR models

The underlying systems have the following general form,

$$\mathbf{z}_t^i = \sum_{k=1}^{K} A_k \mathbf{z}_{t-k}^i + D\mu_t + \mathbf{u}_t, \qquad i \in \{0,1\},\ t = 1..T(\tau), \tag{4}$$

which is a standard VAR model in terms of econometrics. Whether $\mathbf{u}_t$ can be regarded as normally and/or independently distributed can be tested, but we apply resampling methods that do not require normality, and as a robustness check we have also tried a block bootstrap to accommodate serial correlation (without noticeable differences). In order not to overload the notation we do not explicitly differentiate the model parameters according to the model variant i, because the context should make that sufficiently clear. The deterministic part $\mu_\mathbf{t}$ contains a constant term and also seasonal dummies, because it appears that the seasonal adjustment from the statistical agencies did not completely eliminate seasonal patterns. The sample end point $T(\tau)$ is not fixed because in the pseudo-out-of-sample evaluation the estimation sample varies, so τ indexes the shifting evaluation samples. It would be straightforward in practice to let the starting period of the sample move in parallel to the end period to obtain a rolling window.

The h-step point forecasts based on the sample endpoint T are denoted by $\hat{\mathbf{z}}_{T+h|T}^i$ and are computed in a standard recursive fashion:

$$\hat{\mathbf{z}}_{T+2|T}^i = \hat{A}_1 \hat{\mathbf{z}}_{T+1|T}^i + \sum_{k=2}^{K} \hat{A}_k \mathbf{z}_{T+2-k}^i + \hat{D}\mu_{T+2}. \tag{5}$$

Such an iterative multistep forecast is well known to be optimal if the true model is a VAR. If the innovations were normally distributed, the forecast error distribution would also be Gaussian due to the linearity of the system. However, if the residuals follow a non-normal distribution the distribution of the forecast errors would be unknown in general. We use the symbol $\hat{f}_{T+h|T}^i$ for the joint probability distribution function (pdf) of the h-step forecast, or predictive density, with an associated covariance matrix $\Psi_{T+h|T}^i$. The point forecast for the r-th release of industrial production is written as $\hat{y}_{T+h|T,\,r}^{*i}$, and the corresponding marginal density forecast is $\hat{f}_{T+h|T}^i(y_r)$.

Notice that the sequence of forecast errors $\hat{\mathbf{z}}_{T+1|T}^i - \mathbf{z}_{T+1}^i$,..., $\hat{\mathbf{z}}_{T+h|T}^i - \mathbf{z}_{T+h}^i$ will also be correlated as a moving average of the future innovations, and we will need the respective covariance matrix for the industrial production releases in the growth-rate specification, $\Upsilon_{h|T}^0(y_r^*)$.

Having estimated the parameters with the sample ending in $T(\tau)$, note that an h-step forecast corresponds to the reference period $T + h - p - (r-1)$, but this differs between the components of $\hat{\mathbf{z}}^i_{T+h|T}$ because of varying parameters r and p. If $h < p + r - 1$ this is sometimes called a backcast since the reference period of the forecast precedes the current publication period, and the term nowcast is used for $h = p + r - 1$.

The log-level variant with $\mathbf{z}^1_t$ implicitly takes into account the existence of multiple cointegrating relations between variables. The growth-rate variant with $\mathbf{z}^0_t$ neglects the levels information, but this could be negligible in practice. The differencing transformation may provide prediction stability if there are shifts in the level relations (Clements and Hendry, 1999). Ultimately, the choice between the model variants is an empirical question.[16]

The available sample in this model framework is limited by the requirement that the data source must provide the revision history even for the earlier reference periods. In our case the datasets support a starting date of 1993m1 for the US (data from the Alfred database of the St. Louis Fed) and 1995m10 for Germany (mainly from the Bundesbank real-time database).

The coefficients A_k are eliminated, estimated, and used for forecasting as already explained in section 2.2, i.e. they are re-searched and re-estimated in each bootstrap iteration.

3.3. Density forecasts and estimating the turning-point probability

We must distinguish whether the VAR is specified in (log) levels ($\mathbf{z}^1_t$ with $\mathbf{d}_t$) or in growth rates ($\mathbf{z}^0_t$ with $\Delta\mathbf{d}_t$) because the forecast is about cumulative growth over several months. For the specification in growth rates we thus have to combine all predicted one-period growth rates from $T + h - 5$ through $T + h$.

For the log-level specification with $\mathbf{z}^1_t$ we simply need the distribution of the h-step forecast of (the r-th release of) log industrial production. If the innovations are normal, we get standard textbook formulae for the estimated covariance matrix $\hat{\Psi}^1_{T+h|T}$ of the multivariate predictive density, see for example Lütkepohl (2007, e.g. section 3.5). Therefore the following expression would be directly operational:

$$\hat{f}^1_{T+h|T} = N(\hat{\mathbf{z}}^1_{T+h|T},\ \hat{\Psi}^1_{T+h|T}). \tag{6}$$

The forecast error variance of industrial production would be the corresponding element on

[16]Estimating the log-level specification on the non-stationary data without imposing the cointegration restrictions requires a further assumption in order to be able to conduct valid inference about the significance of coefficients in a standard way. Here we rely on the results by Dolado and Lütkepohl (1996) that any levels VAR with at least two lags yields correct t-ratios. For the plausible case of cointegration we can additionally refer to West (1988) showing that usual inference applies. We do not view these requirements as restrictive and assume that they hold, apart from the fact that the preferred US model is the stationary one.

the diagonal of $\hat{\Psi}^1_{T+h|T}$, let us call the square root of this element and thus the standard deviation of this output forecast error $\hat{\psi}^1_{T+h|T}(y^*_r)$. On the other hand, if the innovations cannot be assumed to be jointly normal, the shape of the density $\hat{f}^1_{T+h|T}$, including the marginal distribution of the output forecast $\hat{f}^1_{T+h|T}(y^*_r)$, is unknown in general and we resort to the bootstrap method already explained.

Then we subtract from the distribution of the h-step forecast the (log) value of the corresponding variable six reference months earlier in order to calculate the cumulative growth rate. This starting reference period is $T+h-p-(r-1)-6$, where r and p are variable-specific but known. For this reference period the j-th publication happens in period $T+h-(r-1)-6+(j-1)$, and by definition the latest available release was published in T, which is therefore given by $j^* = \min(6+r-h,\, rmax)$. Thus we get the following expression for the recentered distribution:

$$\hat{f}^1_{T+h|T}(y^*_r) - y_T^{T-(5+r-h-j^*)} \tag{7}$$

For example, if we forecast the value of the second release ($r=2$) three publication months ahead ($h=3$), the reference period is $T+h-p-(r-1)=T$ (a nowcast), the reference period for subtraction is $T-6$, and in principle the latest available release in T for this reference period would be $j=6+r-h=5$. However, $j^* = rmax < 5$ if the higher releases are not available in the dataset.

Finally, we calculate the percentage of how many bootstrap trials fall below the selected threshold which defines a recession. Such a threshold may be interpreted in light of the so-called "triangle approach" by Harding and Pagan (2002), and the choice of the threshold was discussed in section 2.2.3.

If the system is specified in growth rates, an explicit cumulation of forecast errors from step 1 to step h is necessary. The errors of the growth forecasts across horizons are not independent because of the moving-average nature of the multistep forecast, and these correlations have to be taken into account when analyzing the sum of the forecasts. Here we have to resort to simulation methods in any case, irrespective of whether the innovations are normally distributed or not. As before, for horizons $h<6$ the known $6-h$ growth rates of the past are added in order to always consider a uniform 6-month time span. With these simulated frequency distributions of the cumulative growth rate we can then work as described before.

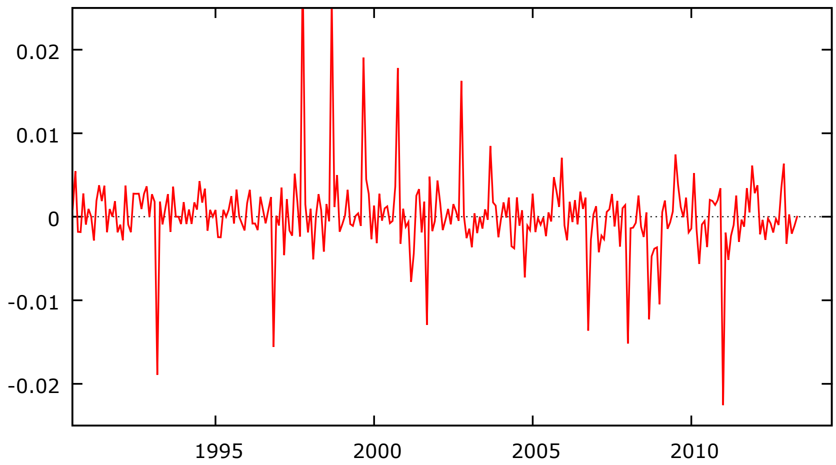

Figure 7: Revisions of US industrial production data. Difference of the second release published in a certain period and the first release published one period earlier, referring to the same reference period (both in logs).

4. Applications of the fully real-time approach

4.1. USA

For the US the main data source for real-time data is the ALFRED database of the St. Louis Fed, from which we always use the vintages that are current at the end of every month, and we re-base the earlier vintages to be comparable to the current margin. The publication lag for US industrial production is only one month, $p = 1$.

Figure 7 reports the most important part of the publication history of the industrial production data, namely the revisions one month after the first release. This distribution (of differences of logs) displays a very high kurtosis, with the bulk of percentage revisions being quite small in absolute value, but a number of noticeable large data revisions. For example, the first raw release for the reference period 1997m10 was 122.6940 published on Nov. 17th, 1997 (with a base period 1992=100), but the release for this observation period which was valid at the end of December, published on Dec. 15th, was changed to 126.3710 (with an unchanged base period). In log-differences these values yield the number 0.02953 which is shown (albeit truncated) in the figure.

We set $rmax = 2$ and thus consider the publications of the first two months after the publication lag of one month has passed. The vectors $\mathbf{z}_t^i$ contain 15 elements, among which the first two releases of industrial production, new orders of durable goods, and of the CPI. The Doornik-Hansen test for normality of the system residuals (for the initial estimation sample) rejects with a p-value of 0 ($\chi^2_{30} = 183$), justifying our non-Gaussian resampling approach.

We observe that in this shorter sample including revision information the fit of the forecast probabilities to the great recession is quite a bit better than in the pseudo real-time simulation in section 2, especially concerning the early months of the recession; see figure 8. For example, the fit of the nowcast of the growth-rate subset VAR (third line in table 5) now attains levels comparable to those of the MS backcast in table 1. This property is also reflected in the detailed timing analysis of the recession signals, see table 6. Running the corresponding model in log-levels revealed that there the estimated probabilities are much more volatile, see online appendix B.

At the bottom of table 5 we have also included the results of a slightly different forecasting approach ("single-equation multi-step"). It consists of an equation for industrial production with the same predictors as for the VAR system. However, instead of calculating the forecasts iteratively, the equations are of the form $y_{T+h} = \sum_k \beta_{k,h} x_T$, where x_T denotes the value of the predictor available at the sample end T, and h is the econometric forecasting step. There are different equations for each forecast step h, with different coefficient estimates. This approach is also sometimes called "direct multi-step" in the literature. Because the first $h-1$ lagged values are not available at T and have to be omitted by construction, serial correlation in the residuals is common and thus we use standard robust covariance estimates for inference. We then apply the same subset restriction search as in the VAR context by restricting insignificant regressors. The density of the forecast is available analytically by standard textbook methods, including parameter uncertainty, where we invoke an asymptotic normality assumption.

4.2. Germany

The main source for the German data is the real-time database at the Bundesbank. In figure 9 we report the realizations of the first revisions (differences of logs). It can be seen that the magnitude of revisions is sometimes substantial. Note also that the first release has been traditionally biased downward because in a simple (unreported) autoregression the constant term turns out as significantly positive, with the coefficients of all autoregressive terms being negative. This may have partly changed very recently from about 2011 onwards as can be seen at the end of the graph, but in general it means that there is systematic information in the revision process that may help the forecast.

We set again $rmax = 2$ and thus only consider the first two publications, which for the given data turned out to capture most of the revision process. The vectors $\mathbf{z}_t^i$ contain 22 elements: first

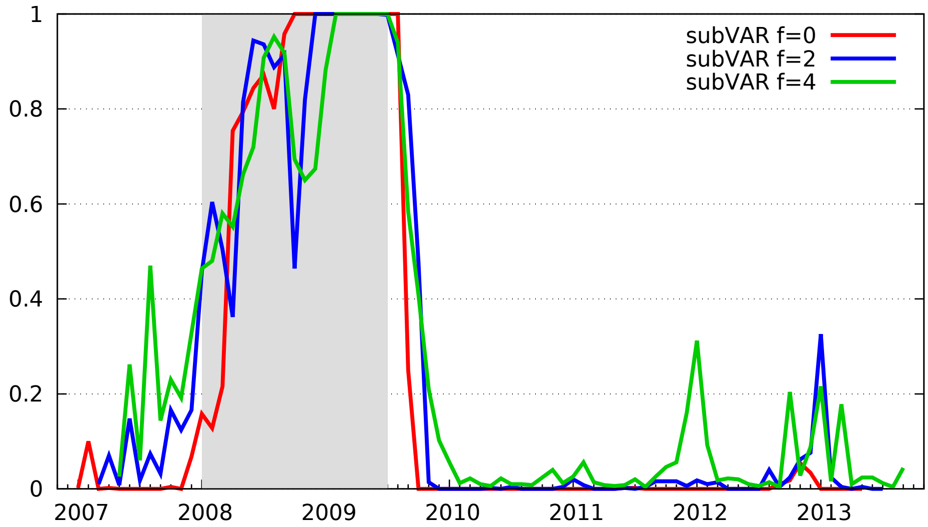

Figure 8: Subset VAR recession probabilities, estimated probabilities of a recession at effective forecast steps $f = 0$ (nowcast) and $f = 2, 4$. The models are specified in growth rates, contain all predictors, real-time data vintages, and the first two data releases where applicable. The time axis denotes reference periods, shaded areas are NBER recession episodes.

Table 5: Evaluation measures of the full real-time simulation, USA

	effective forecast $f = 0$			effective forecast $f = 2$			effective forecast $f = 4$		
	RMSE	MAE	Theil	RMSE	MAE	Theil	RMSE	MAE	Theil
AR(1)	0.312	0.135	0.320	0.365	0.222	0.405	0.410	0.318	0.533
levels subVAR	0.93	0.96	1.00	0.84	0.80	0.81	0.89	0.83	0.71
growth-rate subVAR	*0.76*	*0.58*	*0.80*	*0.66*	*0.47*	*0.63*	*0.60*	*0.44*	*0.50*
single-equation multi-step	0.90	0.77	0.89	0.88	0.63	0.86	1.05	0.85	0.88

Notes: All results relative to the respective AR(1) density forecast (at the top, estimated in log levels). For the US with publication lag 1 the effective forecast step f is one less than the econometric forecast horizon h. The best value across models for a given step f is given in *italics*. For the Theil U1 formula see table 1, and for an explanation of the single-equation multi-step see the text.

Table 6: Signals in the full real-time simulation, US

	effective step f	Signal shift s at peak 2007m12	$s+f$	s at trough 2009m6	$s+f$	Further misclassifications
AR(1)	0	$+2M$	2	$+2M$	2	2007m2, 2008m8, 2011m6/m7,
	2	$+2M$	4	$+1M$	3	2007m2, 2008m6 – 2008m9, 2011m4/m5, 2012m11
	4	$+9M^a$	13	$-3M$	1	2008m11/m12, 2011m3, 2012m9
levels subVAR	0	$+4M^b$	4	$+3M$	3	2008m8
	2	$-3M$	-1	$+2M$	4	2008m1 – 2008m3, 2008m8
	4	$-3M$	1	$-2M$	2	2007m4, 2008m1 – 2008m3, 2009m7 – 2009m9, 2010m7 – 2010m9, 2010m11, 2011m2
growth-rate subVAR	0	$+3M$	3	$+2M$	2	–
	2	$-1M$	1	$+1M$	3	2008m2, 2008m8
	4	$-2M$	2	$-1M$	3	–
single-equation multi-step	0	$+2M$	2	$+2M$	2	2007m2, 2007m4, 2011m6/m7
	2	$+2M$	4	$+2M$	4	2008m8
	4	$+4M$	8	$+1M$	5	2009m10/m11

Notes: Publication dates of estimates 2007m1-2013m5.
Here signals are defined as crossing 50% probability. A leading signal would be published up to (and including) the respective NBER peak/trough dates, and the signal shift s in months is negative for a signal date up to the peak/trough date, zero for a signal immediately after the peak/trough, positive for a lag. The effective forecast step f denotes the difference between the reference period about which the forecast is made and the real-time publication date of the information set.
All models work with density forecasts and a recession definition of 6-month cumulated growth below -1%.
[a] An isolated spike occurs already at $+2M$, but then the following six months would have to be interpreted as misclassifications.
[b] An isolated spike occurs already at $-1M$, but then the following four months would have to be interpreted as misclassifications.

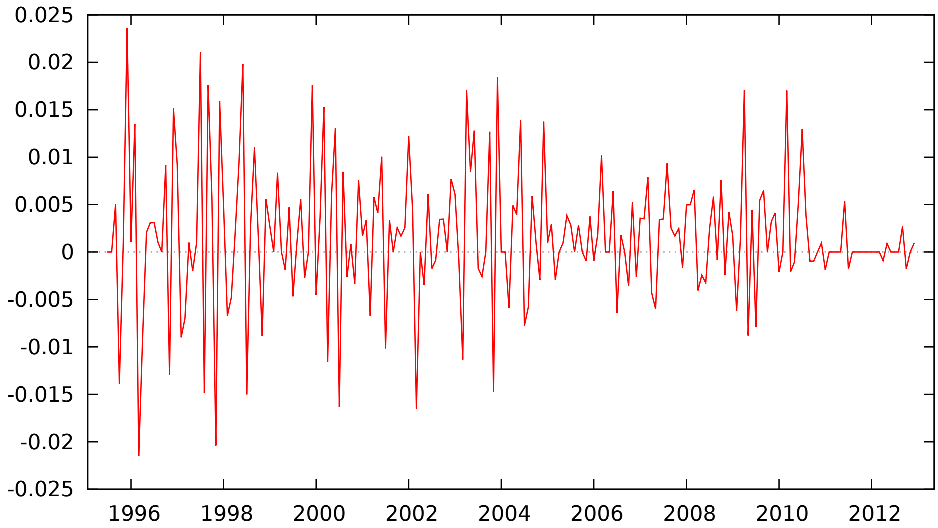

Figure 9: Revisions of German industrial production data. Difference of the second release published in a certain period and the first release published one period earlier (both in logs).

and second publications of the variables industrial production, domestic orders, foreign orders, the (consumer) price index, as well as the variables for which no revision history is available: oil prices, CDAX stock index, REX bond market value index, vacancies, construction permits, the Euribor interest rate, the yield spread for corporate bonds, four different interest rate term spreads with respect to the three-month bond yield, a business climate (IFO) and expectations index (ZEW). Some variables are treated as stationary, namely the interest rate spreads and the Ifo and ZEW indicators.

The maximum possible lag is 4 in the German data sample, and this is also the AIC choice. For the log-levels specification, eliminating all superseded releases in the system implies 252 zero restrictions with four lags which yields a p-value of 0.49. This nominal test result would mean that the information of the revision histories does not significantly contribute to the forecast. However, notice that in a slightly shorter sample up to 2006m1 the p-value of this test changes to 0.028. Furthermore, in a similar system with three lags we got a p-value of 0.027 (for 168 zero restrictions). While this case is not quite as clearcut as for the US, we still view it as an indication of the importance of the revision information for the forecast. The system test for normally distributed residuals again rejects with p-value 0 ($\chi^2_{42} = 109$).

For Germany there exists no established or consensus business-cycle chronology similar to the NBER dating. However, the recent findings of the CEPR Euro Area Business Cycle Dating

Committee (EABCDC, see http://cepr.org/content/euro-area-business-cycle-dating-committee) contain some information on Germany, which we have collected in table 7. It can be seen that those announcements happened with lags of more than one year. A complication is that there is no pronouncement about a German peak between the troughs of 2009Q1 and the one of the euro area in 2013Q1. However, the recent October 2015 findings show (figures 4 and 7) that German GDP had negative quarter-on-quarter growth rates in 2012Q4 and 2013Q1, implying a peak in 2012Q3. The developments of hours worked (figure 6) are slightly ahead with a peak in 2012Q1. For lack of a better solution we simply pick the average date as the peak: 2012Q2.

In figure 10 we display the 6-month growth rate of German industrial production along with the -1% threshold that we use again. Because there are no consensus recession dates we do not perform a calibration of the threshold. However, German adjusted GDP data reveals very lackluster developments in the episodes 1995Q4-1996Q1, 2000Q2-Q4, 2001Q2-2002Q1, 2002Q4-2003Q1, and 2004Q2-2005Q1. This would broadly coincide with the observations when IP growth was below the threshold, except that GDP growth was not very low around 1999Q2/Q3 when IP growth plunged. Nonetheless, the comovement of IP and GDP appears to be weaker in Germany than in the US, making recession forecasts more difficult with monthly data. Historically the German growth rates have been close to zero quite often even in expansion phases, rendering the distinction between expansions and contractions more difficult objectively.

For the following results it is important to bear in mind that the publication lag in Germany is two months (instead of one), i.e. a first publication refers to two months ago. Thus an econometric one-step forecast of the first IP release for Germany means an effective backcast ($f = -1$), a two-step forecast is an effective nowcast ($f = 0$), and so forth.

In terms of the fit of the forecast recession probabilities as well as the timing of the signals it turned out that in the German case the specification in (log) levels performed better. Figure 11 displays the estimated probabilities of the levels system (the results of the system in stationary form are provided in online appendix B), and in table 8 the corresponding evaluation measures of both specifications are shown. It can be seen that the improvement of the subset VAR approach over the related benchmark AR(1) model is much less pronounced than for the US. The detailed timing analysis in table 9 contains many misclassified periods at the strict rule of crossing the 50% mark, but the corresponding plots in figure 11 reveal that many of them are due to probabilities close to 50% in 2008, and high recession probabilities in the year preceding the 2012/13 recession. Nonetheless, we reiterate that capturing recessions in Germany through industrial production data is not quite as successful as in the US context.

Table 7: German recession episodes after 2005

peak	trough	source	peak month	trough month	length in months
2008Q1	2009Q1	CEPR March 2009, Sept. 2010	2008m2	2009m3	13
2012Q2	2013Q1	own interpolation, CEPR Oct. 2015	2012m5	2013m3	10

Notes: CEPR refers to the findings of the Euro Area Business Cycle Dating Committee. The quarterly findings are mechanically translated to monthly periodicity by picking the middle month of the peak quarter and the final month of the trough quarter, because by NBER convention the trough period is regarded as part of the recession. For our interpolation of the peak in 2012 see the text.

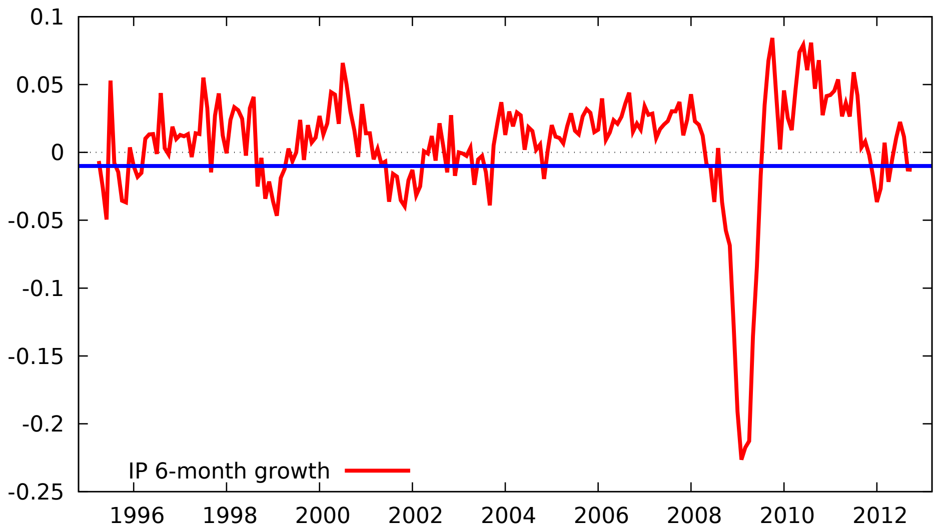

Figure 10: German industrial production 6-month growth. Sample period 1995m4-2012m10, the values are from the respective second releases. The time axis corresponds to the reference period. The straight blue line is at −1% .

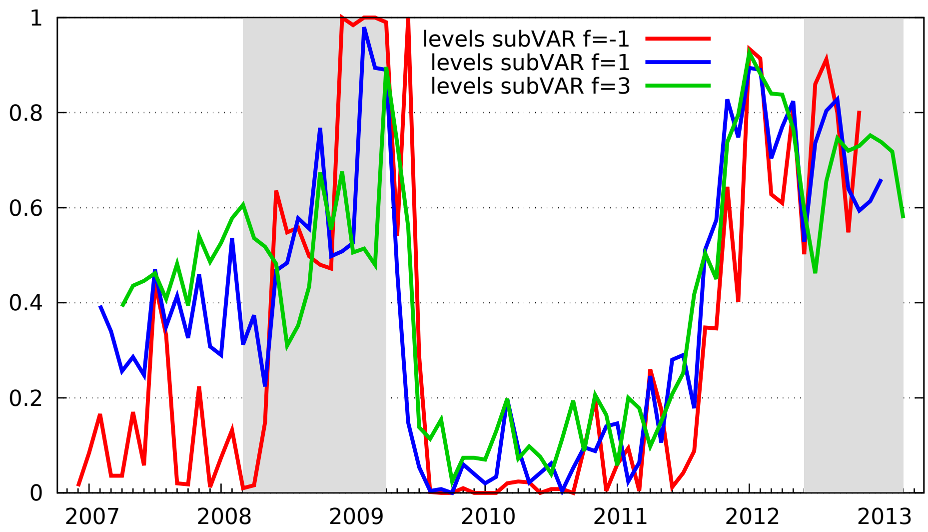

Figure 11: German levels subset VAR recession probabilities, estimated probabilities of a recession at effective forecast steps $f = -1$ (backcast) and $f = 1, 3$. The models are specified in (log) levels, contain all predictors, real-time data vintages, and the first two data releases where applicable. The time axis denotes reference periods, shaded areas refer to the episodes in table 7.

Table 8: Evaluation measures of the full real-time simulation, Germany

	effective forecast $f = -1$			effective forecast $f = 1$			effective forecast $f = 3$		
	RMSE	MAE	Theil	RMSE	MAE	Theil	RMSE	MAE	Theil
AR(1)	0.426	0.247	0.440	0.475	0.375	0.475	0.497	0.460	0.488
levels subVAR	*0.94*	1.05	*0.92*	*0.88*	0.88	*0.86*	*0.90*	*0.83*	*0.86*
growth-rate subVAR	1.03	1.17	1.00	1.03	1.08	0.95	1.04	1.03	0.93
single-equation multi-step	1.04	*0.98*	1.01	0.98	*0.75*	0.90	–	–	–

Notes: All results relative to the respective AR(1) density forecast (at the top, estimated in log levels). For Germany with publication lag 2 the effective forecast step f is two less than the econometric forecast horizon h. The best value across models for a given step f is given in *italics*. For the single-equation multi-step forecast beyond step $h = 4$ ($f = 2$) no predictors remain in the model, thus the bottom right is left blank. (For an explanation of the single-equation multi-step see the text.)
For the Theil U1 formula see table 1.

Table 9: Signals in the full real-time simulation, Germany

	effective step f	Signal shift s at peak 2008m2	$s+f$	s at trough 2009m3	$s+f$	s at peak 2012m5	$s+f$	Further misclassifications with respect to episodes in table 7
AR(1)	−1	$+4M$	3	$+4M$	3	$+6M$	5	2008m10, 2012m2/m3
	1	$+4M$	5	$+3M$	4	$+6M$	5	2007m6, 2008m10, 2010m2, 2011m12 – 2012m3, 2012m9 – 2012m11
	3	$+4M$	7	$\pm 0M$	3	$+5M$	8	2007m6, 2008m8, 2008m10, 2009m12, 2010m2, 2010m11, 2011m1, 2011m8, 2011m11, 2012m2, 2012m7 – 2012m10
levels subVAR	−1	$+4M$	3	$+4M$	3	$-4M$	−5	2008m10 – 2008m12, 2011m12
	1	$+4M$	5	$\pm 0M$	1	$-10M$	−9	2008m1, 2008m10
	3	$-5M^a$	−2	$\pm 0M$	3	$-1M$	2	2007m8, 2008m3 – 2008m6, 2008m12, 2011m6 – 2012m3
growth-rate subVAR	−1	$+4M$	3	$+5M$	4	$-4M$	−5	2011m12, 2012m7 – 2012m11
	1	$+3M$	4	$+5M$	6	$+1M$	2	2007m8, 2007m10, 2008m1, 2010m2, 2011m10 – 2012m4, 2012m9 – 2012m11
	3	$-4M$	−1	$+6M$	9	$\pm 0M$	3	2007m6, 2007m8/m9, 2008m4 – 2008m6, 2009m12, 2010m2, 2011m8 – 2012m2, 2012m9, 2012m11
single-equation multi-step	−1	$+6M$	5	$+5M$	4	$-4M$	−5	2011m11/m12, 2012m5, 2012m7, 2012m9, 2012m11
	1	$+3M$	4	$+2M$	3	$\pm 0M$	1	2007m10, 2008m10, 2011m9 – 2012m4, 2012m9

Notes: Publication dates of estimates 2007m1-2012m12. The CEPR-determined trough occurs after the sample end, in 2013Q1.
Here signals are defined as crossing 50% probability. A leading signal would be published up to (and including) the respective peak/trough dates inferred from table 7, and the signal shift s in months is negative for a signal date up to the peak/trough date, zero for a signal immediately after the peak/trough, positive for a lag. The effective forecast step f denotes the difference between the reference period about which the forecast is made and the real-time publication date of the information set.
All models work with density forecasts and a recession definition of 6-month cumulated growth below −1%.
[a]Borderline signals occur from $-7M$ through $+4M$, accordingly some episode before or after would have to be classified as mismatch.

5. Conclusions

In this paper we have proposed to forecast business-cycle turning points with monthly data using linear systems that fully account for the publication lags and revisions of the data in real time. Our approach uses the forecast probability distribution (predictive density) to infer the probability of a recession. A justified real-time criterion for the existence of a recession was a 6-month cumulated growth rate of industrial production below -1%. Theoretical advantages of our method are efficiency, an implicit modelling of the revision process, and numerical stability. Furthermore, since in our approach the forecasting stage is completely separated from the specification of the interesting regimes, it was shown to be easily possible to implement an arbitrary number or type of exogenously defined (or calibrated) regimes, without having to reconsider the estimation step. For example, the non-recession density region could be differentiated into a stagnation and a genuine expansion area.

The benchmark univariate AR(1) model already incorporated the idea of analyzing the predictive density for the determination of turning points. In our VAR models as well as some considered comparison models we included a relatively broad information set, which appears to be essential for a competitive forecasting model. As a further variant one could combine the general approach with factor models and/or setups that deal with mixed-frequency data (see Schumacher and Breitung, 2008, for Germany). We leave that for future research.

Using US and German data in pseudo and fully real-time out-of-sample forecast evaluations we showed that the turning points can be predicted several months before official data publications confirm them. But of course no miracles should be expected from our approach, either, as it remains difficult to produce informative business-cycle forecasts beyond a horizon of a few months. By using standard evaluation measures as well as the formal model confidence set procedure of Hansen, Lunde, and Nason (2011) we have shown that our suggested method is competitive and sometimes superior, especially under absolute error loss. But in the spirit of Hamilton (2011, abstract) it is our aim to complement other methods, not to replace them: "[T]here may be gains from combining the best features of several different approaches." For example, completely different model types including the present one could be used for forecast model averaging.

Acknowledgements

This paper grew out of a project funded by the German federal ministry of finance, and Schreiber thanks Daniel Detzer for excellent research assistance. For helpful comments we thank (without implicating) Jörg Breitung, Malte Knüppel, Dieter Nautz, Tommaso Proietti, Christian Schumacher, Boriss Siliverstovs, Thomas Theobald, Klaus Wohlrabe, and the late Jürgen Wolters, for discussions at the German Economic Association (VfS) conference 2013, the IAAE and CFE conferences 2014, and at various seminars. We also thank the Soroban HPC cluster at Free University Berlin for computing resources. Special thanks to Christian Proaño and Thomas Theobald for making their forecast results available, and to anonymous reviewers whose suggestions have led to major improvements of this study.

References

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

Brüggemann, R. (2004): *Model Reduction Methods for Vector Autoregressive Processes*, vol. 536 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin.

Chauvet, M., and J. D. Hamilton (2006): “Dating business cycle turning points,” in *Nonlinear time series analysis of business cycles*, ed. by C. Milas, P. Rothman, and D. van Dijk, pp. 1–54. Elsevier, Amsterdam.

Clements, M., and D. F. Hendry (1999): *Forecasting non-stationary economic time series*. MIT Press.

Corradi, V., A. Fernandez, and N. R. Swanson (2009): “Information in the Revision Process of Real-Time Datasets,” *Journal of Business & Economic Statistics*, 27(4), 455–467.

Cottrell, A., and R. Lucchetti (2013): *Gretl User's Guide*Version 1.9.14; available at http://gretl.sourceforge.net/#man.

Dolado, J. J., and H. Lütkepohl (1996): “Making Wald tests work for cointegrated VAR systems,” *Econometric Reviews*, 15(4), 369–386.

Fresoli, D., E. Ruiz, and L. Pascual (2015): “Bootstrap multi-step forecasts of non-Gaussian VAR models,” *International Journal of Forecasting*, 31, 834–848.

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

Hansen, P. R., A. Lunde, and J. M. Nason (2011): “The Model Confidence Set,” *Econometrica*, 79(2), 453–497.

Harding, D., and A. Pagan (2002): “Dissecting the cycle: A methodological investigation,” *Journal of Monetary Economics*, 49(2), 365–381.

Kauppi, H., and P. Saikkonen (2008): "Predicting U.S. recessions with dynamic binary response models," *The Review of Economics and Statistics*, 90(4), 777–791.

Kilian, L. (1998): "Accounting for Lag Order Uncertainty in Autoregressions: the Endogenous Lag Order Bootstrap Algorithm," *Journal of Time Series Analysis*, 19(5), 531–548.

——— (2001): "Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order," *Journal of Forecasting*, 20, 161–179.

Krolzig, H. M. (1997): *Markov switching vector autoregressions. Statistical inference and application to business cycle analysis.* Springer Science+Business Media.

Lütkepohl, H. (2007): *New Introduction to Multiple Time Series Analysis.* Springer.

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.

Ng, E. C. Y. (2012): "Forecasting US recessions with various risk factors and dynamic probit models," *Journal of Macroeconomics*, 34(1), 11–125.

Nyberg, H. (2010): "Dynamic probit models and financial variables in recession forecasting," *Journal of Forecasting*, 29(1-2), 215–230.

Österholm, P. (2012): "The limited usefulness of macroeconomic Bayesian VARs when estimating the probability of a US recession," *Journal of Macroeconomics*, 34, 76–86.

Proaño, C., and T. Theobald (2014): "Predicting Recessions with a Composite Real-Time Dynamic Probit Model," *International Journal of Forecasting*, 30(4), 898–917.

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(3), 386–398.

Stock, J. H., and M. W. Watson (1993): "A procedure for predicting recessions with leading indicators: Econometric issues and recent experience," in *Business Cycles, Indicators and Forecasting.* J. Stock and M. Watson, eds., NBER Book Series Studies in Business Cycles.

Theobald, T. (2014): "Markov Switching with Endogenous Number of Regimes and Leading Indicators in a Real-Time Business Cycle Application," *The Empirical Economics Letters*, forthcoming.

Villani, M. (2009): "Steady-state priors for vector autoregressions," *Journal of Applied Econometrics*, 24(4), 630–650.

West, K. D. (1988): "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica*, 56(6), 1397–1417.