Freie Universität Berlin

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

Sven Schreiber

School of Business & Economics
Discussion Paper

Economics

2014/2

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

by Sven Schreiber†

this version: January 2014

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 the relevant underlying variables, and to derive the probability of the turning point from the forecast density as the probability mass below (or above) a given 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. Then we apply the method to US and German monthly data. In an out-of-sample exercise (for 2007-2012/13) the turning points can be signalled before the official data publication confirms them (but not before they happened in reality).

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

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

*This paper grew out of a project funded by the German federal ministry of finance, and we thank Daniel Detzer for excellent research assistance. For helpful comments we thank (without implicating) Jörg Breitung, Malte Knüppel, Dieter Nautz, Christian Schumacher, Boriss Siliverstovs, Thomas Theobald, Klaus Wohlrabe, and Jürgen Wolters, for discussions at the German Economic Association (VfS) conference 2013 and at other seminars.
The empirical applications including the data processing were carried out with the *gretl* econometrics software, see Cottrell and Lucchetti (2013), especially chapter 8 on real-time data.

†Macroeconomic Policy Institute (IMK) Düsseldorf, and Free University Berlin (Institute of Statistics and Econometrics). Email: svetosch@gmx.net, postal address: Hans-Böckler-Str. 39, D-40476 Düsseldorf, Germany.

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. 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.[1] The general idea is that we use a linear (system) model to predict the continuous output variable several steps 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 five months to call a recession.[2]

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.[3] 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. Finally, because of the linearity of the estimator the method is computationally robust. In contrast, for the typically applied non-linear model classes it is always possible that the maximization of the likelihood function could fail numerically.

There are also drawbacks of our approach which have to be acknowledged: In order to

[1]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).

[2]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.. Often two consecutive quarters (i.e., six months) are implicitly considered as a minimum recession duration. However, not all "official" recessions fulfill this criterion, and thus we chose five months instead to shorten our turning point recognition lag.

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

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.[4]

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 get 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 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), and working with the predictive densities. However, our conclusions are quite different from his pessimistic outlook, since our models perform quite well. The main difference appears to be the size of the model, as

[4]With respect to the alternative of a Bayesian estimation see the discussion in section 2.4.

[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.

Österholm focuses on trivariate systems, and thus our findings highlight the importance of a broad information set. For a further discussion of Bayesian methods in this context see section 2.4.

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 Nyberg (2010), Ng (2012), and Proaño and Theobald (2012). The main alternative is Hamilton's Markov-switching approach, and some recent applications with real-time data are Hamilton (2011), Nalewaik (2012), or Theobald (2012).

In the rest of the paper, the following section formally presents the statistical approach, including the notational conventions for the real-time data with publication lags and revisions, and (in subsection 2.5) we explain how we construct and use the density forecasts to infer the recession and expansion probabilities. Next, we present the data of the US and German applications. Afterwards the in-sample properties of the estimated models are presented. Then, in section 4.3 we present the pseudo out-of-sample results for turning-point / recession probability forecasts, which cover the period of the recent "great recession", before we conclude.

2 The statistical real-time framework

2.1 Notation

It is not so common to use VAR models for real-time data with publication lags and possible revisions, and thus we introduce some useful definitions and notation to capture these properties. As always in the real-time literature we must distinguish two time axes: first the time period when a value is published, and secondly the time period when the measured activities or events actually took place. Throughout, we call the former "publication period" and the latter the "reference period". A well-known data presentation or storage 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"). Our time granularity is monthly in both dimensions, so in order to simplify matters we ignore any intermediate publications that are superseded again after only a couple of days.

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

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

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 the first result of their measurement. The publication lag is of course different across variables, but we assume that it is fixed over time for each variable.[6] Without loss of generality 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.

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 representa-

[6] 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.

tion is as follows.

- 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 which represent different time series and where even 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 is meant to still use 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 propose instead to also consider the other diagonals of the triangular real-time data matrix representation, in addition to the main diagonal. For each economic variable x we define the following collection $x_{t,1}^*, x_{t,2}^*, ..., x_{t,rmax}^*$ of statistical variables:

Table 1: The real-time notation in relation to the triangular-matrix representation

publication periods $\rightarrow$ $\downarrow$ 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 $p = 1$.

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 value *rmax* represents a cutoff point which is arbitrarily chosen to limit the number of regressors later in the model, but in theory the revision process can go on forever. An example of the correspondence between the different notations and the triangular-matrix representation is given in table 1.

The notable feature of this new data representation is that each statistical variable now only has a single time index instead of two, and this time index also defines the available information set. Therefore the values can be written as elements of standard time series vectors and a standard lag operation can be unambiguously defined on the single time index: $Lx_{r,t}^* = x_{r,t-1}^* = x_{t-1}^{t-1-p-(r-1)}$. It is this representation which enables us to use standard VAR tools.

Another advantage relative to the traditional real-time approach that has focused on the columns of the triangular matrix is that now 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 also used. 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 $p = 2$). This means that we have implicitly also a model of the revision process itself and exploit any systematic movements in there for our forecasts.

A final piece of notation is the mapping $t()$ between the human-readable labels of time periods such as "2010m4" and the integer-valued time period index.

The publication lag p will actually be zero for many variables, especially for financial-market data which are immediately available. The maximum publication lag in our application will be 2 periods (months), for example concerning the German industrial production index.

For any considered variable $x^*_{r,t}$, if it is stationary it is an element of the (column) vector $\mathbf{s}_t$ of dimension $n_0 \times 1$, or if it is integrated it belongs in another vector $\mathbf{d}_t$ of dimension $n_1 \times 1$,[7] where each revision-prone variable x will count as $rmax$ statistical variables in n_0 or n_1. Since the revision process of the data should not introduce any additional non-stationary behavior, 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:

$$\begin{aligned} \mathbf{z}_t^0 &= (\mathbf{s}_t', \Delta\mathbf{d}_t')' \\ \mathbf{z}_t^1 &= (\mathbf{s}_t', \mathbf{d}_t')' \end{aligned} \tag{2}$$

In the model using $\mathbf{z}_t^0$ the integrated variables are thus differenced to render all variables stationary, whereas the vector $\mathbf{z}_t^1$ is a mixture of stationary and integrated variables in levels. Both vectors have dimension $(n_0 + n_1) \times 1$.

Because the topic of this paper is the detection of turning points of economic activity, we are ultimately interested only in the output variable (measured by log industrial production

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

in our monthly data), which we denote with y; for the US data the publication lag here is one month, and in the German dataset two months. Since (log) output is non-stationary, the collection $y^*_{1,t},...,y^*_{rmax,t}$ is a part of $\mathbf{d}_t$.

2.2 The statistical VAR models

The underlying two systems for each country have the following general form,

$$\mathbf{z}^i_t = \sum_{k=1}^{K} A_k \mathbf{z}^i_{t-k} + D\mu_t + \mathbf{u}_t, \qquad i \in \{0,1\},\, t = 1..T(\tau), \tag{3}$$

which is a completely standard VAR model in terms of econometrics, with multivariate white-noise residuals. Whether $\mathbf{u}_t$ can be regarded as normally distributed remains to be tested. 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. We refer to $i = 1$ as the (log-) level specification, and to $i = 0$ as the growth rate specification. The deterministic part $\mu_\mathbf{t}$ contains a constant term and also centered 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 here because we conduct a pseudo-out-of-sample evaluation, where the end of the estimation sample varies. This means that the systems are re-estimated for every publication period in the evaluation sample; an alternative approach could be to hold the parameter estimates from an earlier sample as fixed and make use of newer data only for the forecasts. We could also easily consider a rolling window for estimation by letting the starting period of the sample move in parallel to the end period; this may be desirable in practical on-going applications of our method if structural breaks are suspected in the beginning of the sample. In our formal presentation in this paper, however, we assume a time-invariant data-generating process (DGP).

The only purpose of the systems (3) and of the subset variants described below is to produce h-step forecasts. The point forecasts based on the sample endpoint T are denoted by $\hat{\mathbf{z}}^i_{T+h|T}$ and are computed in a standard recursive fashion, e.g.

$$\hat{\mathbf{z}}^i_{T+2|T} = \hat{A}_1 \hat{\mathbf{z}}^i_{T+1|T} + \sum_{k=2}^{K} \hat{A}_k \mathbf{z}^i_{T+2-k} + \hat{D}\mu_{T+2}. \quad (4)$$

Such an iterative multistep forecast is well known to be optimal if the true model is a VAR. Of course the future realizations of the variables are random due to the innovations $\mathbf{u}_{T+1}, ..., \mathbf{u}_{T+h}$. 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 $\hat{f}^i_{T+h|T}$ for the joint (multivariate) probability distribution function (pdf) of the h-step forecast, or predictive density, with an associated covariance matrix $\Psi^i_{T+h|T}$. The point forecast for the r-th release of industrial production is written as $\hat{y}^{*i}_{T+h|T,r}$, and the corresponding marginal density forecast is $\hat{f}^i_{T+h|T}(y_r)$.

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

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 using $\mathbf{z}^1_t$ implicitly takes into account the existence of multiple cointegrating relations between variables. The second model variant with $\mathbf{z}^0_t$ by construction includes only stationary variables, because such specifications are often used for short-term forecasts of inflation in the literature. Here the levels information is neglected, i.e. the implicit error correction terms of the level model are not included, but this could be negligible for relatively short time horizons. The differencing transformation of the variables may provide a certain stability of the predictions if there are shifts in the level relations (Clements and Hendry, 1999). Ultimately, the choice between the model variants is an empirical question.

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.

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

2.3 Sequential model reduction

Our estimation procedure for a given sample end period τ begins with the usual OLS estimates of each equation of this system, where we choose the lag parameter K based on the best $\bar{R}^2$ in the levels equation for the first publication of industrial production. The next steps are the automatic sequential elimination of insignificant regressors to improve the overall precision of the estimation.[8] The regressor with the highest p-value of the usual two-sided t-statistic is removed and the equation is re-estimated (again with OLS).

[8] For computational efficiency we perform this elimination separately for each equation; it would be straightforward in principle to conduct it on the system level with the seemingly unrelated regressions (SUR) method, which is not necessarily superior in practice, however. For a comprehensive discussion see Brüggemann (2004).

This procedure is repeated until no t-statistic has a p-value above the selected threshold p_{cutoff}, where we have set $p_{cutoff} = 0.05$.

Once all the equations have been reduced in this way, the entire system is re-estimated with the SUR method, i.e. in an efficient manner, taking into account the cross-equation covariance matrix of the residuals. This feasible-GLS estimator of course also has analytical solutions and thus never fails. The final reduced system then provides the basis for the real-time forecasts based on the current margin τ.

We denote the final SUR-estimated coefficients with the imposed zero restrictions on some elements through the described sequential model reduction by $\hat{B}_k$. The resulting estimated subset-VAR model is thus:

$$\mathbf{z}_t^i = \sum_{k=1}^{K} \hat{B}_k \mathbf{z}_{t-k}^i + \hat{D}\mu_t + \hat{\mathbf{u}}_t, \qquad i \in \{0,1\},\ t = 1..T(\tau), \tag{5}$$

with a corresponding standard estimate of the cross-equation residual covariance matrix in this sample, $\hat{\Omega}_T$.

2.4 Discussion of alternative solutions to the curse of dimensionality

The problem of scarcity of degrees of freedom in large dynamic models can be solved in different ways, all of which are a kind of "shrinkage" estimation methods. Firstly, the dimension of the data space can be reduced by the extraction of common factors or principal components. These factor models are an active field of research and application, for a recent application to German real-time data see Schumacher and Breitung (2008). However, there it is not easy to determine the relevance of certain predictors. Alternatively one can reduce the parameter space: For example, if there is a priori information on the coefficients, Bayesian methods can be applied. Recently Banbura, Giannone, and Reichlin (2010) showed that big BVARs can have competitive prediction qualities, and with a suitable choice of priors there exists no initial degree-of-freedom problem. However, it is often dubious whether there actually is a priori information about the coefficients; rather it seems that BVARs are simply used in the literature as a convenient shrinkage tool. In the latter case, a more classical approach in the spirit of the general-to-specific methodology is also possible by successively removing the insignificant predictors in order to increase the efficiency of the forecast. This subset VAR approach with automated zero restrictions is what we use in this paper.

2.5 Density forecasts and estimating the turning-point probability

In the following we explain in some detail how to infer regime probabilities from the real-time VAR forecasts. While for the sake of clarity we focus on recession forecasts, the approach can easily be adapted to focus on the probability mass in the upper part of the forecast density, relating to an expansion regime. In our empirical application below, we will analyze both types of turning points.

In the following sections we must distinguish whether the VAR is specified in (log) levels (with $\mathbf{d}_t$) or in growth rates (with $\Delta\mathbf{d}_t$). The reason is that for the specification in growth rates we need to combine all predicted growth rates from $T+h-4$ through $T+h$ by cumulation to arrive at the total growth.

Note that for the estimation of the forecast density we consider only the innovation uncertainty and ignore the additional estimation uncertainty of the parameters. However, this should be important mostly in the tails of the distributions, and these are the areas which are not very relevant for our purposes. After all, we are not (very) interested whether a recession probability is really 6% instead of 2%, but we care more about whether the probability is 30% or 70%. Nevertheless, in principle it would be possible to take parameter uncertainty into account, too; although the needed nested bootstrap simulations would increase the computational burden considerably.

2.5.1 Level specification

The log-level specification with $\mathbf{z}_t^1$ is the simpler case, as 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 under normal innovations:

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

The forecast error variance of industrial production would be the corresponding element on 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 a straightforward bootstrap method: We draw from the estimated residuals with replacement and simulate 500 forecast paths over the h-step horizon. (Since in our application the center of the distribution is more relevant than the margins, we can work with this relatively low number of repetitions.) We then simply use this empirical distribution as approximating the true predictive density sufficiently. Of course, this simulation must be re-run for each new observation that is added to the sample.

Now we subtract from the distribution of a h-step forecast the (log) value of the corresponding variable five reference periods (months) earlier in order to calculate the cumulative growth rate over this five-month period. This benchmark reference period is $T+h-p-(r-1)-5$, where r and p are still given from the variable which is being forecast. For this reference period the i-th publication happens in period $T+h-(r-1)-5+(i-1)$, and we always use the latest available release in real time published in T, which is therefore given by $i^* = \min(5+r-h, rmax)$. Thus we get the following expression for the recentered distribution:

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

The complicated time and release indexing reflects the fact that we need to pick data that refers to an actual five-month lag with respect to the economic activity, not with respect to the date when the latest available release was published. 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 $T - 5$, and in principle the latest available release in T for this reference period would be $i = 5 + r - h = 4$. However, $i^* = rmax < 4$ if the higher releases are not available in the dataset. Another example: forecasting the initial publication ($r = 1$) five publication months ahead ($h = 5$) implies that we must also use the initial publication ($i^* = 1$) for subtraction which was published in period T, because the later releases would not have been published yet from a real-time perspective.

The intermediate result that we obtain is the density forecast of the predicted cumulative five-month growth rate for this h-step horizon, and the variance is unchanged since the subtraction of a known value only re-centers the distribution. Finally, we determine how much probability mass falls 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), where for simplicity we use a fixed 5-month period for evaluation, and our critical value is $-0.025 * 2/5 = -1\%$ which is an arbitrary but commonly used value.[9]

Therefore we want to determine –under normal innovations– the CDF value of the normal distribution with mean $\Delta_5 \hat{y}^{*1}_{T+h|T,r} = \hat{y}^{*1}_{T+h|T,r} - y_T^{T-(5+r-h-i^*)}$ and standard deviation $\hat{\psi}^1_{T+h|T}(y^*_r)$ at the given value of -1%. This of course is equivalent to finding the CDF value of the standard normal distribution denoted with $\Phi()$, at the corresponding stan-

[9] For another recent example of comparing a predictive density with an exogenous threshold value see Galbraith and van Norden (2012).

dardized value:

$$\Phi\left(\left(-0.01-\Delta_5\hat{y}^{*1}_{T+h|T,r}\right)/\hat{\psi}^{1}_{T+h|T}(y^*_r)\right), \tag{8}$$

This standard numerical problem is the only non-linear step in our approach. This value gives the probability that the publication of the r-th release in period $T+h$ will meet the definition of a recession.

Under non-normal innovations, we simply calculate the percentage of the re-centered simulated forecasts which fell below the threshold value.

2.5.2 Specification in growth rates

If the system is specified in growth rates, an explicit cumulation of forecast errors from step 1 to step h is necessary, in parallel to the fact that we are interested in the cumulated (or average) growth over this period. The determination of the distribution of the cumulative growth rate is complicated by the fact that the errors of the growth forecasts across horizons are not independent because of the moving-average nature of the multistep forecast. We called the covariance matrix of this h-dimensional distribution $\Upsilon^0_{h|T}(y^*_r)$, and these correlations have to be taken into account when analyzing the sum of the forecasts.

Since the shape of such a composite forecast density appears to be yet unknown in the literature, here we have to resort to simulation methods in any case, irrespective of whether the innovations are normally distributed or not. However, while we use a resampling scheme under non-normal innovations as in the levels case, we employ a parametric Monte Carlo approach if the residuals are normally distributed, based on their estimated covariance matrix $\hat{\Omega}_T$ to draw random realizations of this multivariate (normal) distribution. We again abstract from the parameter uncertainty, as in the levels case.

As before we focus especially on output (log industrial production) up to forecasting horizon h. For each simulated path the growth rates are summed to approximate the cumulative growth. For horizons $h<5$ the known $5-h$ growth rates of the past are added in order to always consider a uniform 5-month time span. With these simulated frequency distributions of the cumulative growth rate we can then work as described before.

3 Application to the US

We now apply our proposed method to actual data, beginning with 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.

3.1 Basic data properties

In figure 1 we present the cumulated five-month growth rates of industrial production for each publication month (i.e. the time index in the subscripts in our notation), along with a horizontal line showing the -1% threshold. Apart from some borderline months we have two clear industrial-sector recessions in our complete sample, one between early 2001 and early 2002, and of course the recent Great Recession from the beginning of 2008 until mid-2009. It is also apparent that in the industrial sector episodes of temporary contractions with 5-month growth rates slightly below zero are quite common, and thus it makes sense to use a negative threshold to call a recession. Notice that we use a uniform exogenous recession threshold of -1% over 5 months for both of the analyzed countries, but other definitions of a recession could of course also be used.

Since the central topic of this paper is the real-time perspective, we report in figure 2 the most important part of the publication history of the industrial production data, namely the revisions after one month after the first release, which we simply call the "first revisions". 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

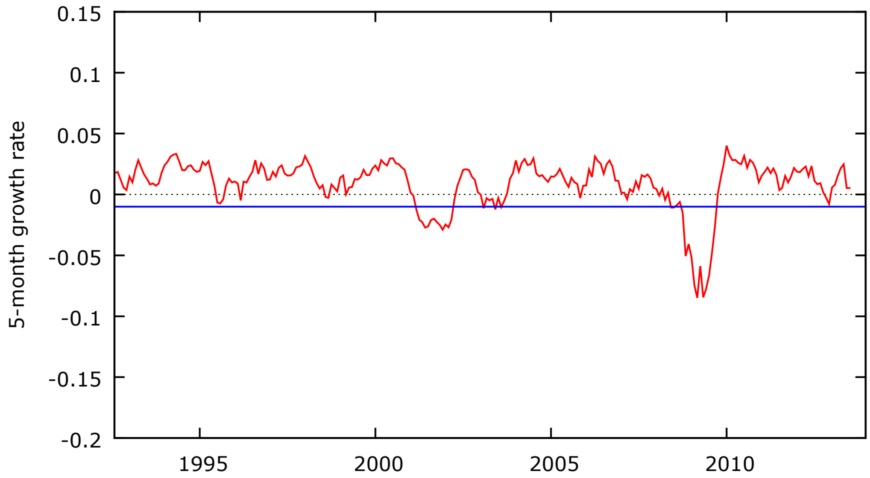

Figure 1: US industrial production, five-month rolling (cumulated) growth. The time axis refers to the period of data publication, i.e. two months after the reference period for this series. The straight blue line is at -1%.

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 only consider the publications of the first two months after the publication lag has passed (which is one month for US industrial production). The vectors $\mathbf{z}_t^i$ contain 15 elements: First two releases of the data with revision information: industrial production, new orders of durable goods, and of the CPI. Furthermore we include oil prices, a stock index, real loans (volume), the conference board composite index, a corporate bond yield spread, the 3-months T-bill rate, which are all non-stationary, and as stationary variables the Michigan confidence index, number of building permits, and a term spread.

3.2 In-sample properties

We chose $K = 4$ based on the best $\bar{R}^2$ in the levels equation for the first release of industrial production. For the specification with growth rates this translates to $K = 3$ because obviously there the fourth lagged level appears implicitly in the differences.

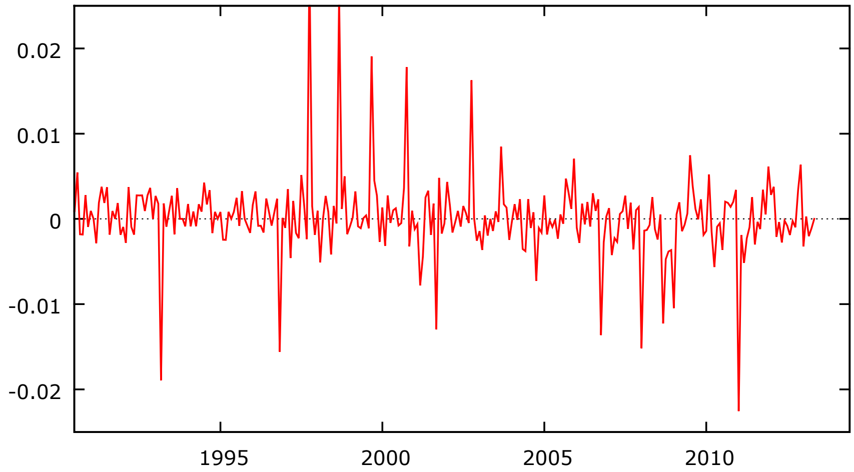

Figure 2: Revisions of US industrial production data. Difference of the second release published in a certain period and the first release published one period earlier, referring to the same reference period (both in logs).

Before applying the subset reduction search in the system, we perform the mentioned test of the null hypothesis that the dynamics of the revision process are not systematically relevant. Remember that this test is only needed in the log-level specification; in the growth-rate specification we expect systematic autocorrelation of revisions in any case due to the differencing, even if revisions of the source data in levels were white noise. We can implement the test for absence of systematic autocorrelation in the levels system as exclusion restrictions on the superseded releases in the system (except for the initial lag, all other lagged first-release series are superseded by later releases). In our four-lag system this corresponds to 135 restrictions, and the standard Wald test rejects the null hypothesis with a p-value of $p = 0.0096$. Therefore we conclude that capturing the dynamics of the revision process is potentially important for forecasting.

Due to the large dimension of the underlying system, it is hardly possible to characterize the properties of all the estimated equations in detail. The figures 3 and 4 show the in-sample forecast errors (one step) of the equations for the industrial production of the two model variants. The variance of the forecast error for the second publication (first revision) is much lower, which is of course due to the fact that the existing information of the first publication can be used to predict the second publication for the same reference

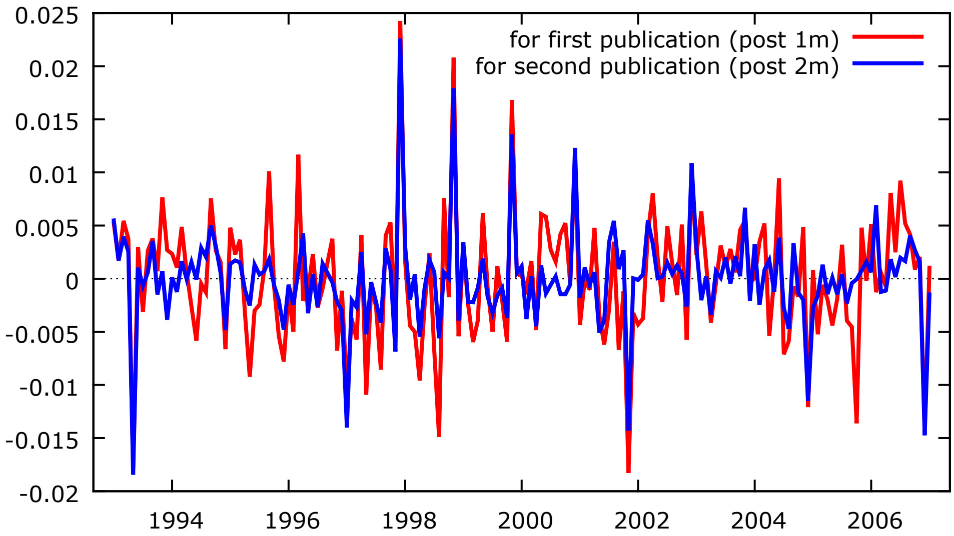

Figure 3: One-step forecast errors (in-sample residuals) of industrial production for the log level specification, US model

Table 2: Test of normality of in-sample forecast errors, US models

	1st release IP equ.	2nd release IP equ.	full SUR-system
log level spec.	13.13 (0.001)	67.84 (0)	441.4 (0)
growth rate spec.	0.56 (0.757)	2.30 (0.317)	567.4 (0)

Notes: Jarque-Bera-type test with $\chi^2(2)$ distribution, or Doornik-Hansen test for system residuals, p-values in parentheses. The forecast errors of the industrial production (IP) equations are displayed in figures 3 and 4.

period, which eliminates a significant portion of the uncertainty.

As described before, in order to derive the predictive density correctly, it matters whether the residuals (one-step forecast errors) can be regarded as normally distributed. In table 2 we report the formal tests of this null hypothesis, with the result that it is mostly rejected, except for the tests based on the single equations in the growth-rate systems. This means that we have to resort to bootstrap methods in any case.

It is of course also interesting which indicators are relevant for the forecast equations, and we give an overview of the equations for industrial production in tables 3 and 4, where we do not distinguish between different lags. Due to the prior automatic variable selection procedures all remaining terms are obviously highly significant by construction,

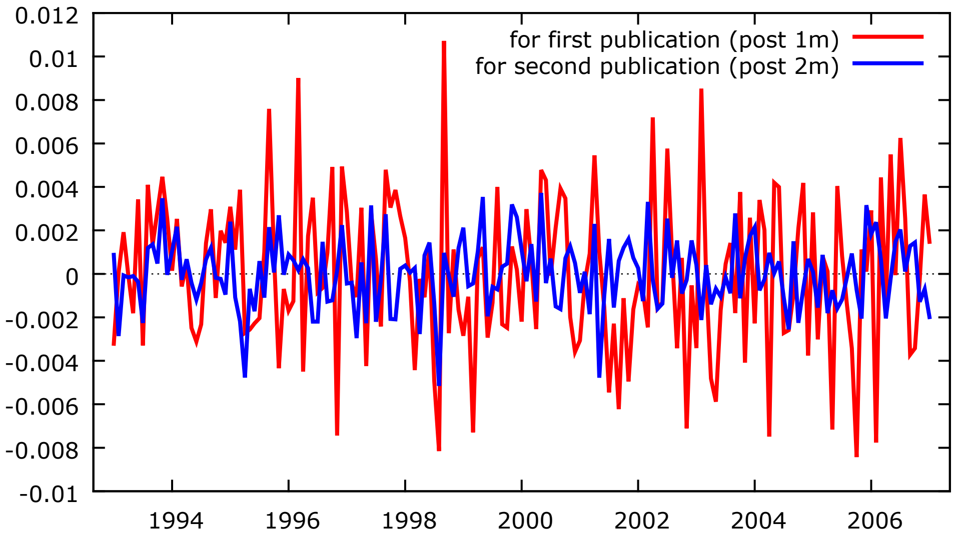

Figure 4: One-step forecast errors (in-sample residuals) of industrial production for the growth rate specification, US model

so we also abstain from reporting p-values. When interpreting this table it should be noted that the removal of the predictor in an equation is only directly relevant for the one-step forecast; for multi-step forecasts of a variable x the predictor z can nevertheless be relevant due to the cross-connections in the system, even if in the equation of x it is not included directly.

The oil price and the number of construction permits never appear directly in the output equations. Somewhat surprisingly, the term spread is also absent at least in the levels specification. In contrast, (lags of) incoming orders and industrial production itself have an important role always. The other indicators are significant in some of the samples and equations.

3.3 Out-of-sample evaluation of the turning point forecasts

We now turn to the forecasting performance of our approach within the available period of evaluation. As already mentioned, the time axis in the following diagrams reflects the real-time information set, i.e. the date of publication; for example at date 2007m1 all data published up to that date are used for estimation. For the case of a one-step forecast,

Table 3: Regressors in the industrial production equations, log-level specification, US

	1st release		2nd release	
sample until	$\tau = 2007m1$	$\tau = 2013m05$	$\tau = 2007m1$	$\tau = 2013m05$
Ind. prod.	x (1r/2r)	x (1r/2r)	x (1r)	x (1r)
CPI		x (2r)		x (1r)
Orders	x (1r/2r)	x (1r/2r)	x (1r/2r)	x (1r/2r)
Term spread				
Conf. board index	x	x	x	x
Michigan index	x			
Oil price				
Stock index	x	x	x	x
Real loans	x	x	x	x
Construct. permits				
Int. rate 3m T-Bill		x		
Corp. spread	x	x	x	

Notes: "1r", "2r" means 1st or 2nd release of the variable as regressor.

Table 4: Regressors in the industrial production equations, growth-rate specification, US

	1st release		2nd release	
sample until	$\tau = 2007m1$	$\tau = 2013m05$	$\tau = 2007m1$	$\tau = 2013m05$
Ind.prod.	x (1r)	x (1r)	x (1r)	x (1r)
CPI	x(1r)	x (1r/2r)		
Orders	x (1r)		x (1r)	x (1r)
Term spread	x	x		
Conf. board index	x	x		x
Michigan index	x			
Oil price				
Stock index	x	x		x
Real loans		x		
Construct. permits				
Int. rate 3m T-Bill	x			x
Corp. spread				x

Notes: "1r", "2r" means 1st or 2nd release of the variable as regressor.

because of the publication lag of one month for the US this implies a prediction of the first release for the current reference month and is thus an effective nowcast; a two-step forecast would be a nowcast when the second release is involved, and so forth. It is crucial that the recession or boom signals (or lack thereof) can be interpreted as real-time signals, whereas the implicitly affected reference months must always be derived using the meta knowledge of the publication lags.

Since our aim is to predict reality, we focus on the forecasts of the final data release that is included in our information set, not on the earlier publications that will become obsolete. In our case that means the second release. On the other hand, targeting later releases also increases the difficulty of the forecast and introduces a kind of additional delay; the question is whether it will still be possible to anticipate turning points. We report the immediate one-step forecast results, the results for the three-step horizon, and finally the longest analyzed five-step horizon.

Figure 5 shows the estimated recession probabilities for the growth-rate as well as log-level specification. The circle symbols at 0 or 1 for each publication period are for comparison, to distinguish between more and less informative signals: if the definition of a recession is already fulfilled ex post by the published data (based on the latest available five-month period) without requiring a forecast, the circle is drawn on top at the value 1. Due to the test results reported above we have treated the innovations as non-normally distributed and thus have estimated the forecast densities with non-parametric simulations.

The exercise is successful in so far as the predicted recession probability rose considerably in 2008 before published data confirmed the fact. For the growth-rate specification this phenomenon is most pronounced in the 5-step forecast, but in the log-level variant the 5-step forecast gave a false signal in early 2007 while it did not anticipate the 2008 recession as early as the 3-step forecast. The exact timing of the recession signal of course also depends on the definition of the signal as crossing a certain probability threshold, which need not be 50%.[10] But in general it must be acknowledged that the ability to fore-

[10] In fact, if a probability threshold of $p = 0.5$ were adopted to define the signal, and under the additional

cast business-cycle turning points appears to be limited to at most one quarter ahead (in publication release time, which is less in terms of reference periods by the amount of the publication lag).

The upswing in late 2009 is also correctly anticipated by the estimated recession probabilities, and for the remainder of the evaluation sample without any further recessionary episodes the estimated probabilities do not increase much anymore. However, as all of these probability series appear quite volatile, we also report their simple (cross forecast horizon) average in figure 6. Using these smoother average probabilities may be desirable for practical purposes, the general idea to combine forecasts from different models to improve the reliability is described in Timmermann (2006).

In order to assess the forecast quality of the different specifications we also compare the results in the evaluation sample quantitatively. Table 5 contains the usual evaluation statistics root mean squared error (RMSE), mean absolute error (MAE), and the Theil measure, for our models as well as for certain competing benchmark models: a univariate AR(1) of the second release of industrial production, a deterministic model with a fitted linear trend, and a so-called "direct multistep forecast" model. This latter model is obtained by simply regressing the $t+h$ values of the industrial production release on the information set in t, yielding a single-equation model for an h-step forecast, where the forecast density is essentially given by the estimated residuals. Reassuringly, the subset-VAR approach is superior to these generic competitors. (Of course the difficulty of the forecast depends on the horizon, and thus the columns of the table are not directly comparable.) The log-level specification tends to be preferable for shorter forecasting horizons, but the differences between the VAR variants are not large in any case.

Finally, with our method we can also characterize the business-cycle outlook in more elaborate ways. This is an advantage of our approach that separates the problem of forecasting from the problem of defining and specifying regimes, contrary to probit or

assumption of a symmetric forecast error distribution it would be sufficient to compute the point forecast and to compare that to the chosen exogenous recession threshold.

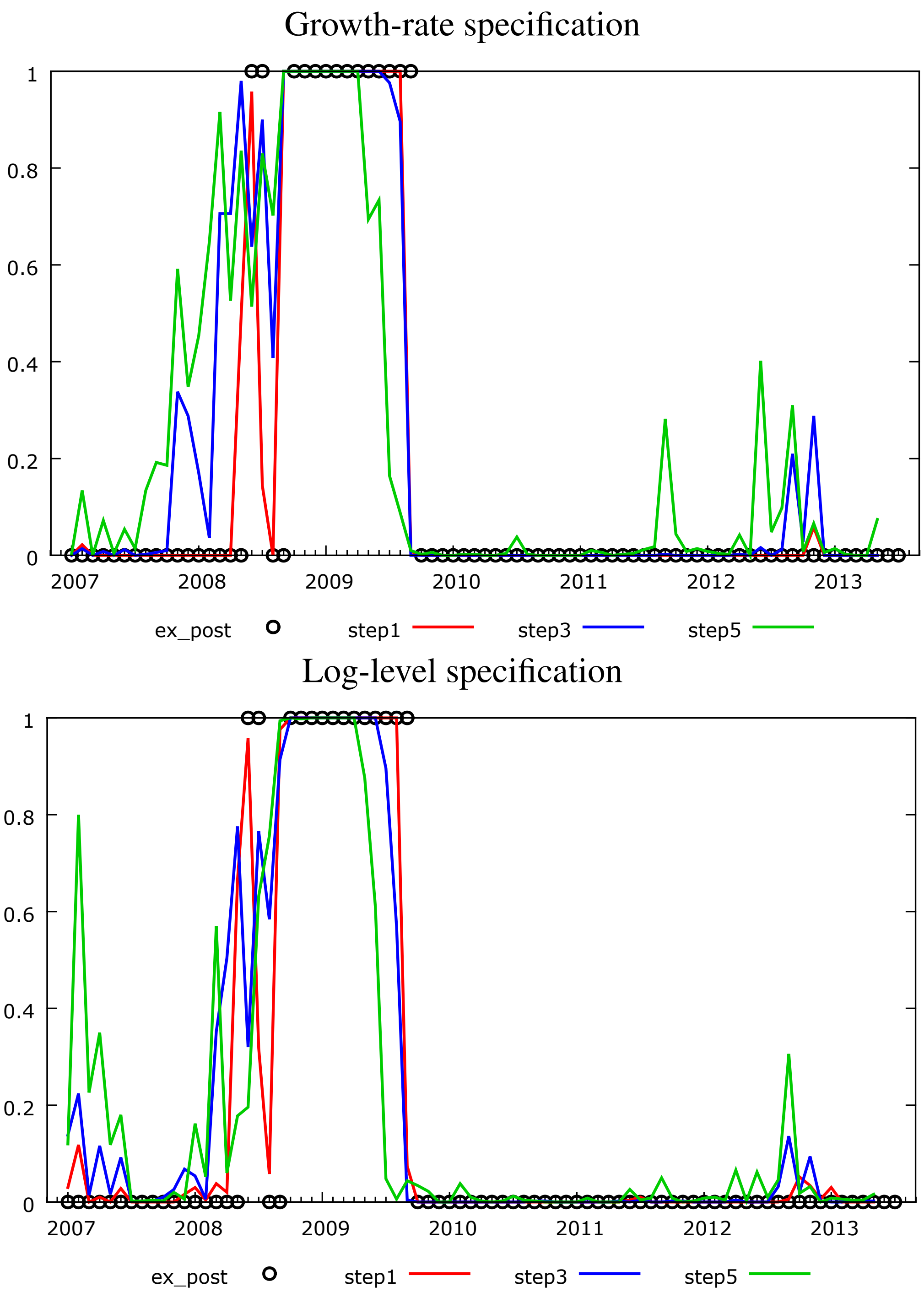

Figure 5: US out-of-sample recession probabilities, referring to the second release. Upper panel VAR specification in growth rates (for the $I(1)$ variables) with at most 3 lags; lower panel specification in log-levels with at most 4 lags. Densities derived with resampling from estimated residuals (see text). Circles represent the binary indicator series of whether the data published in a certain period yielded a 5-month growth rate below the recession threshold ("ex_post").

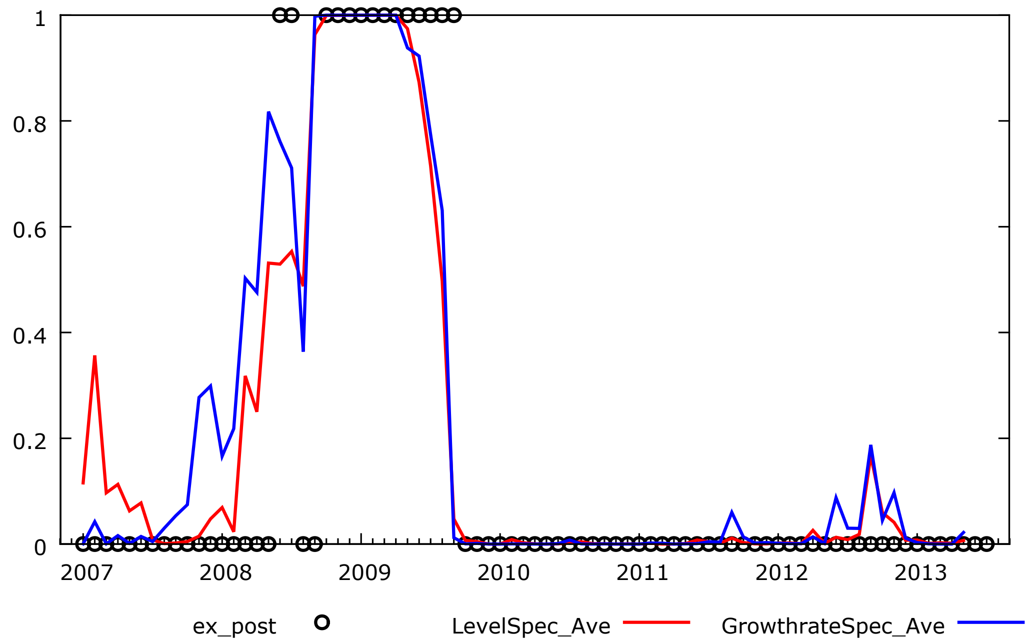

Figure 6: US average out-of-sample recession probabilities – averages of the previously shown forecasting horizons, always referring to the second release.

Table 5: US quantitative out-of-sample forecast evaluation

	1-step			3-step			5-step		
	RMSE	MAE	Theil	RMSE	MAE	Theil	RMSE	MAE	Theil
AR(1)	0.200	0.074	0.247	0.294	0.181	0.391	0.416	0.321	0.651
Linear trend	2.13	3.09	3.49	1.46	1.25	2.22	1.13	0.91	0.95
Log-level VAR	*0.28*	0.23	*0.27*	*0.66*	*0.40*	*0.59*	0.66	0.37	0.51
Growth-rate VAR	0.30	*0.14*	0.29	0.78	0.45	0.67	*0.55*	*0.35*	*0.41*
Direct multistep level-equ.(*)	0.37	0.27	0.36	0.88	0.65	0.79	1.66	2.01	0.88

Notes: All other models' results are expressed relative to the AR(1) results. The comparison refers to the binary realizations (the indicator series of whether the recession definition was met or not, ex post). Forecast of the second release (two months after the reference period in the US case); evaluated against ex-post realizations 2007-2013; the best results are *emphasized*.
*: with lagorder $\geq$ horizon

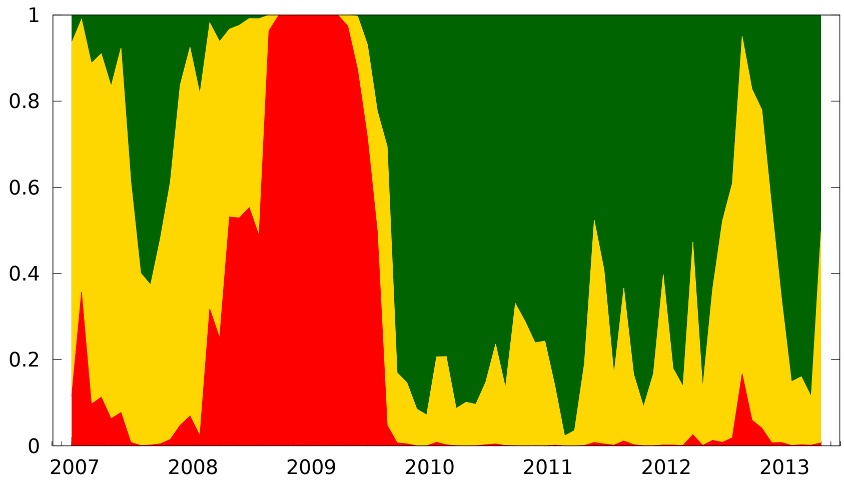

Figure 7: US three-regime probability plot (recession red, expansion green, in-between yellow); underlying are the cross-horizon averages of the log-levels specification.

Markov-Switching methods. A natural extension is to extend our analysis to three situations instead of two; we split the non-recession region into a genuine "expansion" region and an intermediate "stagnation" region, where we fix the additional threshold again exogenously, at a five-month growth rate of 1.2%. Obviously, we do not need to re-estimate anything, we must only calculate another set of quantiles from the estimated densities. The resulting three-region stacked probability plot is displayed in figure 7. We can see that even though the earlier shown recession probabilities were low in 2007, early 2008, and late 2012, this more differentiated analysis reveals that the stagnation probabilities were high, and by implication, that the chances for a noticeable expansion were relatively slim.

4 Application to Germany

Next, we apply our methodology to a German realtime dataset, and we proceed in the same way as in the US case. The main source for the data is the real-time database at the Bundesbank.

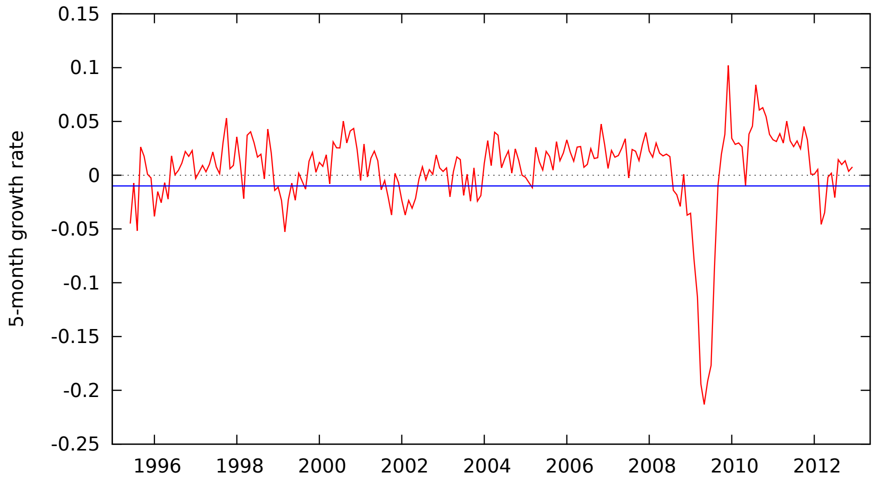

Figure 8: German industrial production, five-month moving average (cumulated) growth. The time axis refers to the period of data publication, i.e. three months after the reference period for this second-release series. The straight blue line is at -1%.

4.1 Basic data properties

In figure 8 we again present the cumulated five-month growth rates of industrial production for each publication month, and it can be seen that in Germany the growth rates of industrial production are more volatile compared to the US, both in the very short und medium runs. The recession threshold of -1% is crossed quite often, producing at least five recessionary episodes in this shorter sample. Thus we expect that for German data it will be comparatively difficult to forecast the state of the business cycle.

In figure 9 we again 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

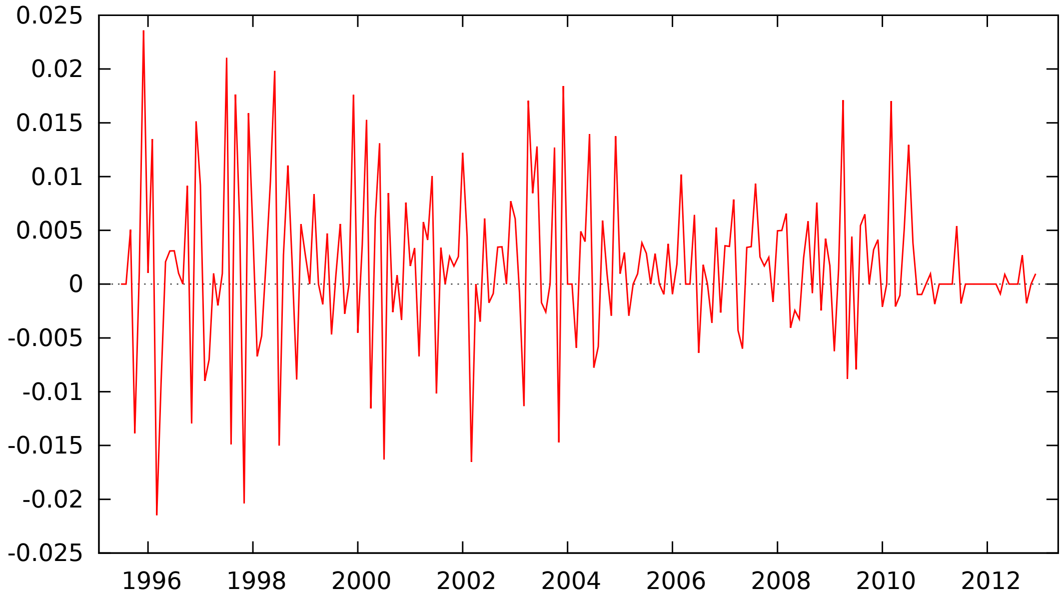

Figure 9: Revisions of industrial production data. Difference of the second release published in a certain period and the first release published one period earlier (both in logs).

elements: first 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). Again variables are appropriately transformed; depending on the model variant, the non-stationary variables are either included in log levels or transformed by log-differencing to growth rates. In any case, some variables are treated as stationary, namely the interest rate spreads and the Ifo and ZEW indicators.

4.2 In-sample properties

It happens that again the lag order $K = 4$ turns out to be preferred. ($K = 3$ for the specification with growth rates.) With respect to the automatic model reduction, in this case our method sets at least half of the coefficients in the system to zero.

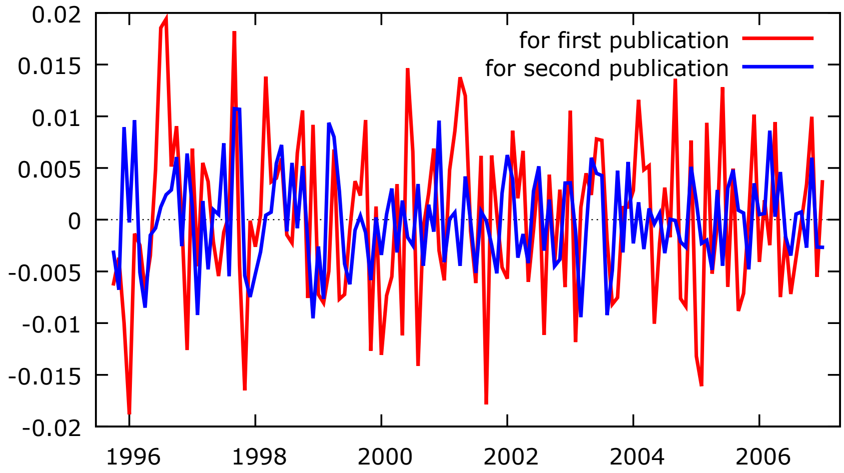

Figure 10: One-step forecast errors (in-sample residuals) for the log level specification

For the log-levels specification with the German data, eliminating all superseded releases in the system implies 252 zero restrictions with four lags which yields a p-value of p = 0.49. This nominal test result would mean that the information of the revision histories does not significantly contribute to the forecast. However, since there are almost twice as many restrictions in the full VAR system compared to the US system, we suspected that many insignificant terms could hide the importance of a subset of significant influences. Thus for the sake of this test we also estimated the levels system with only three lags and re-ran the revision exclusion test; here we got a p-value of p = 0.027 for 168 zero restrictions. While this evidence is not quite as clearcut as in the US case, we still view it as an indication of the importance of the revision information for the forecast.

Next, the figures 10 and 11 show the in-sample forecast errors (one step) of the equations for the industrial production of the two model variants.

In table 6 we report the tests of the null hypothesis of normally distributed innovations, and here this hypothesis typically cannot be rejected based on the univariate residuals of the industrial production equations. (The growth rate specification implies rejection of normality with respect to the first publication, but that is obviously due to some outliers at the end of 1995, which can also be seen in figure 11.) However, the system test overwhelmingly rejects as in the US case, and since our multistep forecast is based on the

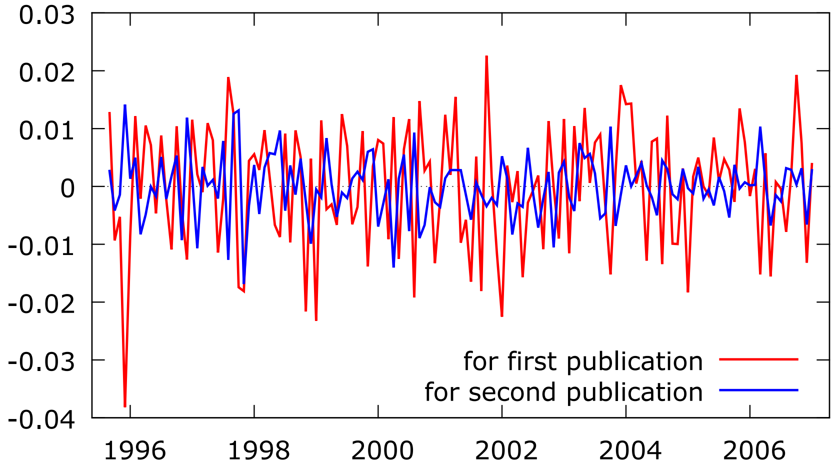

Figure 11: One-step forecast errors (in-sample residuals) for the growth rate specification

Table 6: Test of normality of in-sample forecast errors, Germany

	1st release IP equ.	2nd release IP equ.	full SUR-system
log level spec.	0.26 (0.879)	1.86 (0.395)	220.4 (0)
growth rate spec.	7.50 (0.024)	3.33 (0.189)	241.5 (0)

Notes: Jarque-Bera-type test with $\chi^2(2)$ distribution, or Doornik-Hansen test for system residuals, p-values in parentheses. (Some of) The forecast errors are displayed in figures 10 and 11.

entire system, this may be an argument in favor of simulation methods again.

We proceed with the tables 7 and 8 which give an overview of the non-excluded regressors in the equations for the industrial production. The bottom line in the log-level specification table 7 is that in principle most variables contribute to the overall information. Apart from this it is interesting that the popular Ifo business climate index is only helping the forecast of the first publication, not for further revised releases. In the following table 8 with respect to the specification in growth rates, the oil price is not needed anymore.

4.3 Out-of-sample evaluation

We now evaluate the forecasting performance of our method for Germany, in analogy to the US case. For the case of a 2-step forecast, because of the publication lag of two

Table 7: Regressors in the industrial production equations, log-level spec., Germany

	1st release		2nd release	
sample until	$\tau = 2007m1$	$\tau = 2012m12$	$\tau = 2007m1$	$\tau = 2012m12$
Ind. prod.	x (1r)	x (1r/2r)	x (1r/2r)	x (1r/2r)
CPI	x (1r)	x (1r/2r)	x (1r/2r)	
Foreign Orders	x (1r/2r)	x (2r)	x (1r)	x (1r)
Domestic Orders	x (1r/2r)	x (1r/2r)	x (1r/2r)	x (1r/2r)
Spreads	x (1y/2y/3y/10y)	x (1y/2y/3y/10y)	x (1y/2y/3y/10y)	x (3y)
Ifo index	x	x		
ZEW index	x	x	x	
oil price			x	x
CDAX (stocks)	x	x		
REX (bonds)	x	x		x
vacancies	x	x	x	x
construct.permits	x			
Euribor int.rate	x	x	x	
Corp. spread	x			

Notes: "1r", "2r" means 1st or 2nd release of the variable, "1y" the one-year interest term spread and so on.

Table 8: Regressors in the production equations, growth rate specifications, Germany

	1st release		2nd release	
sample until	$\tau = 2007m1$	$\tau = 2012m12$	$\tau = 2007m1$	$\tau = 2012m12$
Ind.prod.	x (1r/2r)	x (1r/2r)	x (1r/2r)	x (1r/2r)
CPI			x (1r/2r)	
Foreign Orders	x (1r)	x (1r/2r)	x (2r)	x (1r)
Domestic Orders	x (1r/2r)	x (1r/2r)	x (1r)	x (1r/2r)
Spreads	x (1y/2y/3y/10y)	x (1y/2y/3y)	x (2y)	x (1y/2y/3y)
Ifo index		x		
ZEW index			x	x
oil price				
CDAX (stocks)	x	x		
REX (bonds)	x			x
vacancies	x	x	x	x
construct. perm.				x
Euribor int.rate	x	x		x
Corp. spread			x	x

Notes: "1r", "2r" means 1st or 2nd release of the variable, "1y" the one-year interest term spread and so on.

months this implies a prediction of the first publication for the current reference month and thus an effective nowcast.

In figure 12 the upper graph refers to the specification in growth rates, and the lower graph is with respect to the levels specification. The ex-post recession signal based on published data first applies in the (publication) period 2008m8. It is also apparent that at the end of 2008 (in November) there is an occasion where the cumulative 5-month growth lies above the defined recession threshold. The path of the published industrial production data around the end of 2008 thus resembles a double-dip recession pattern, which is repeated to some extent in early/mid 2012.

In general, it is apparent that some predicted recession probabilities are capable of correctly signalling the recessions that (shortly) afterwards appears as a fact in the published data, both in 2008 and 2012. Nevertheless, some forecast variants also falsely indicate recessions that did not happen. Since this appears to be due to the volatility of various predictions we again also consider the average probability over all forecast horizons. The result is shown in figure 13 and is quite encouraging. In particular, the level specification now no longer gets any recession wrong. For the 2008 recession the signal happens just slightly before the ex-post data based signal (in 2008m7), whereas for the second recession 2012 the forecast performance is more favorable: the ex-post data based signal happens in 2012m3, compared to the forecast recession signal in 2011m11. The end of the 2008/2009 recession is signalled two months earlier by the forecast than by the published data (2009m7 vs. 2009m9).

Next, we turn again to quantitative evaluation measures which we report in table 9. The log-level VAR specification has some advantage especially for the 3-step horizon, although the growth-rate VAR specifications performs best at the one-step horizon.

Finally, we also apply the three-situation analysis with an expansion and a stagnation region in addition to recession probabilites. These results are shown in figure 14, and for the German case they show that the stagnation region appears relatively less important

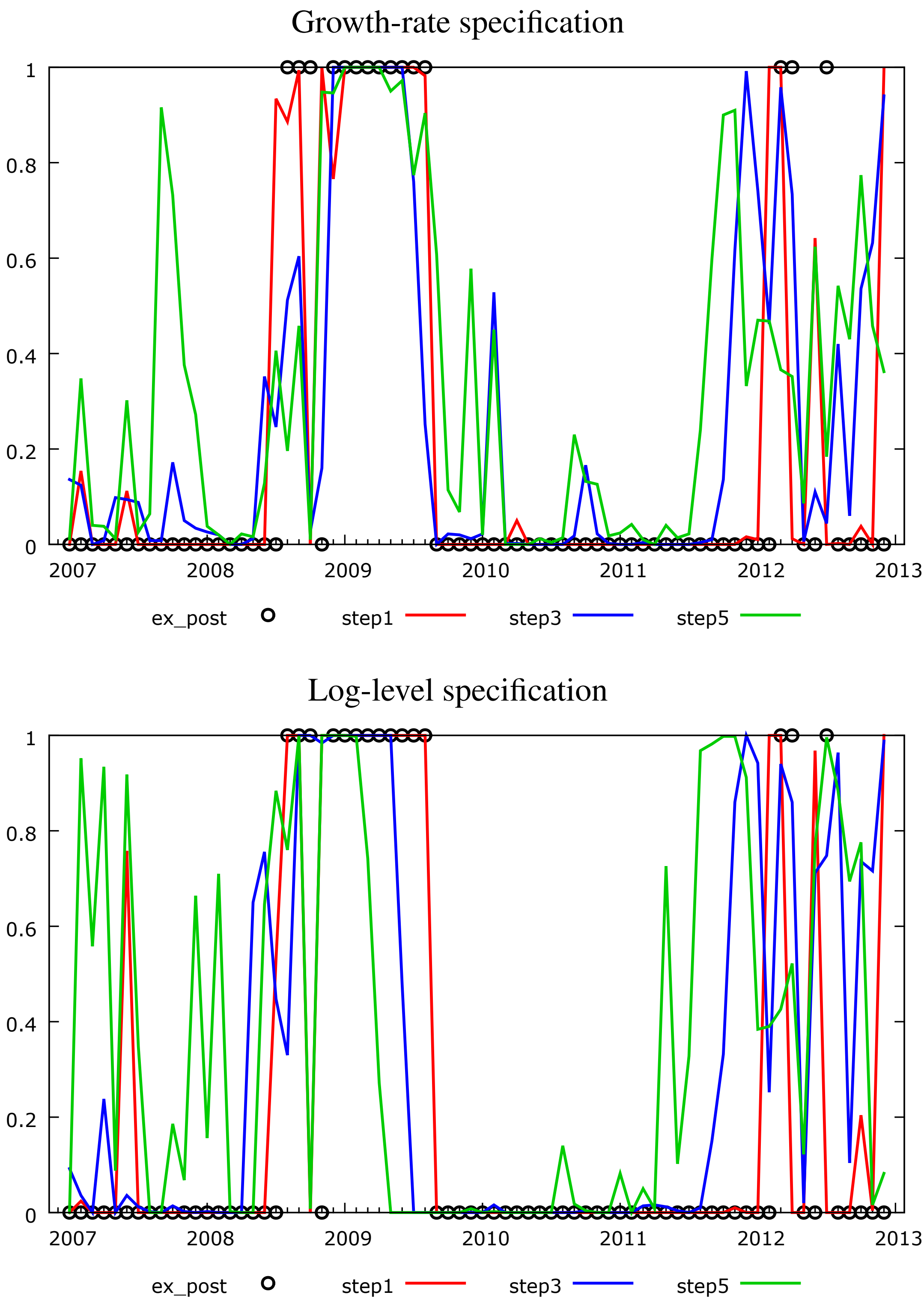

Figure 12: German out-of-sample recession probabilities (rel. 2) – i.e. referring to the second release. Upper panel VAR specification in growth rates (for the $I(1)$ variables) with at most 3 lags; lower panel specification in log-levels with at most 4 lags. Densities derived with resampling of estimated residuals.

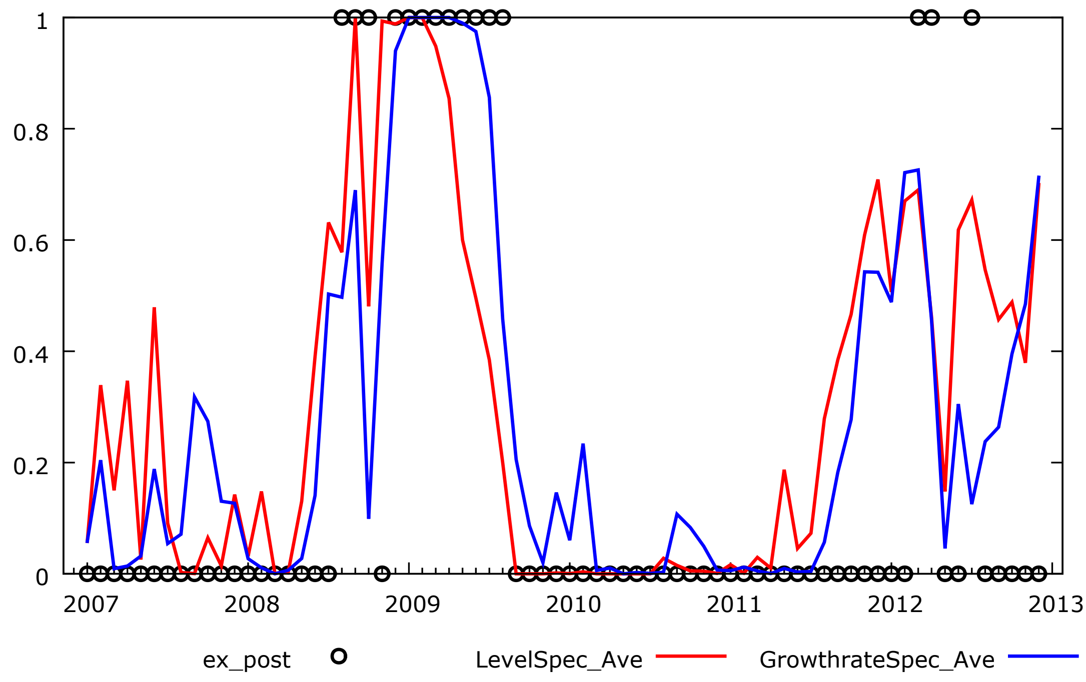

Figure 13: Average out-of-sample recession probabilities (release 2) – averages of the previously shown forecasting horizons, always referring to the second publication.

Table 9: German out-of-sample forecast evaluation

	1-step			3-step			5-step		
	RMSE	MAE	Theil	RMSE	MAE	Theil	RMSE	MAE	Theil
AR(1)	0.263	0.140	0.296	0.354	0.288	0.393	*0.457*	0.435	0.533
Linear trend	1.98	2.92	1.73	1.48	1.41	1.30	1.13	0.91	0.95
Log-level VAR	0.60	0.24	0.57	*0.73*	*0.38*	*0.68*	1.02	*0.66*	*0.87*
Growth-rate VAR	*0.49*	*0.22*	*0.48*	0.97	0.62	0.97	1.01	0.71	0.90
Direct multistep level-equ.(*)	0.58	0.29	0.55	0.92	0.54	0.83	1.59	1.44	1.03

Notes: All other models' results expressed relative to the AR(1) results. The comparison refers to the binary realizations (the indicator series of whether the recession definition was met or not, ex post). Forecast of the second release (three months after the reference period in the German case); evaluated against ex-post realizations 2007-2012; best results are *emphasized*;
*: with lagorder $\geq$ horizon

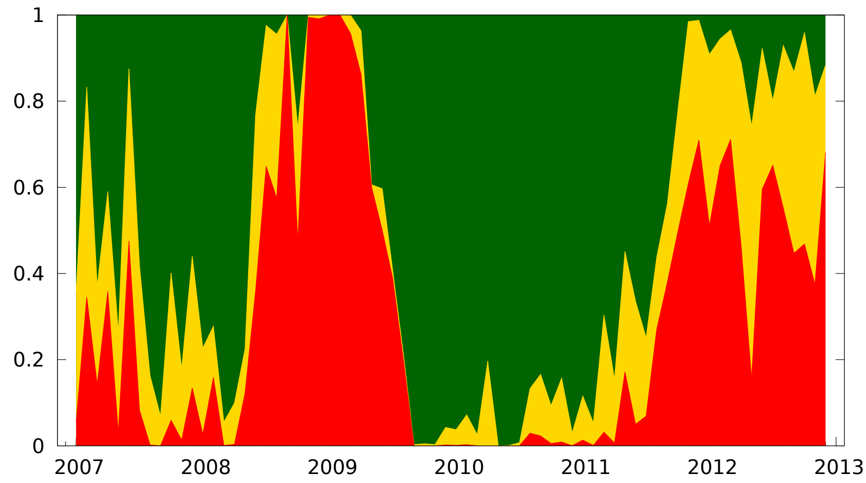

Figure 14: German three-regime probability plot (recession red, expansion green, in-between yellow); underlying are the cross-horizon averages of the log-levels specification.

than for the US, except in 2012.

5 Conclusions

In this paper we have proposed to forecast business-cycle turning points with linear systems that fully account for the publication lags and revisions of the data in real time. Our approach uses the whole forecast probability distribution (predictive density) to infer the probability of a recession. The main advantages over other popular methods are efficiency (if one is prepared to treat the data-generating process as approximately linear), 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 is easily possible to implement extensions such as an arbitrary number of exogenously defined regimes, without having to reconsider the estimation step.

We included a relatively broad information set, which we view as essential for achieving a competitive forecasting model. Indeed, in principle all considered variables appear to be relevant, including timely available unrevised financial-market as well as real revision-prone variables with a publication lag (such as production and new orders), and also sentiment indicators.

Using US and German monthly data and the sample 2007-2012/13 for forecast evaluation purposes, we show that the turning points can be predicted one to several months before official data publications confirm them. Despite the inherent recession recognition lag of any turning-point forecasting method an anticipating signal is possible. 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.

In our view the method suggested here is competitive, 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."

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.

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.

GALBRAITH, J. W., AND S. VAN NORDEN (2012): "Assessing gross domestic product and inflation probability forecasts derived from Bank of England fan charts," *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 175(3), 713–727.

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

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.

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. (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 (2012): "Predicting German Recessions with a Composite Real-Time Dynamic Probit Indicator," Working Paper 5, The New School for Social Research.

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. (2012): "Real-time Markov Switching and Leading Indicators in Times of the Financial Crisis," Working Paper 98, IMK.

TIMMERMANN, A. (2006): "Forecast combinations," in *Handbook of Economic Forecasting*, Vol. 1, pp. 135–196. G. Elliot, C. Granger and A. Timmermann, eds., North-Holland, Amsterdam.

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

Diskussionsbeiträge - Fachbereich Wirtschaftswissenschaft - Freie Universität Berlin
Discussion Paper - School of Business and Economics - Freie Universität Berlin

2014 bereits erschienen:

2014/1 FOSSEN, Frank M.; Ronny FREIER und Thorsten MARTIN
Race to the debt trap? : Spatial econometric evidence on debt in German municipalities
Economics