IMK
Macroeconomic Policy Institute

The IMK is an institute of the Hans-Böckler-Stiftung

WORKING PAPER

No. 188 · December 2017 · Hans-Böckler-Stiftung

EVALUATING THE PREDICTING POWER OF ORDERED PROBIT MODELS FOR MULTIPLE BUSINESS CYCLE PHASES IN THE U.S. AND JAPAN

December 6, 2017

Christian R. Proaño[a] and Artur Tarassow[b]

ABSTRACT

We investigate the probability forecasting performance of a three-regime dynamic ordered probit model framework suitable to forecast recessions, low growth periods and accelerations for the U.S. and Japan. In a first step, we apply a non-parametric dating algorithm for the identification of these three phases. We compare the pseudo-out-of-sample forecasting skills of an otherwise standard binary dynamic probit model with a three-regime dynamic ordered probit framework by means of a rolling-window exercise combined with time-varying indicator selection. Based on a set of monthly macroeconomic and financial leading indicators, the results show the superiority of the ordered probit framework to forecast all three business cycle phases up to six months ahead under real-time conditions. Apart from standard probability forecast evaluation measures, receiver-operating curves and related summarizing statistics are computed.

[a] University of Bamberg, Germany and Centre for Applied Macroeconomic Analysis (CAMA), Australia.
[b] University of Hamburg.

Evaluating the Predicting Power of Ordered Probit Models for Multiple Business Cycle Phases in the U.S. and Japan

Christian R. Proaño[a] and Artur Tarassow* [b]

[a]University of Bamberg, Germany and Centre for Applied Macroeconomic Analysis (CAMA), Australia

[b]University of Hamburg

December 6, 2017

Abstract

We investigate the probability forecasting performance of a three-regime dynamic ordered probit model framework suitable to forecast recessions, low growth periods and accelerations for the U.S. and Japan. In a first step, we apply a non-parametric dating algorithm for the identification of these three phases. We compare the pseudo-out-of-sample forecasting skills of an otherwise standard binary dynamic probit model with a three-regime dynamic ordered probit framework by means of a rolling-window exercise combined with time-varying indicator selection. Based on a set of monthly macroeconomic and financial leading indicators, the results show the superiority of the ordered probit framework to forecast all three business cycle phases up to six months ahead under real-time conditions. Apart from standard probability forecast evaluation measures, receiver-operating curves and related summarizing statistics are computed.

JEL Classifications: C52, C53, C37.

Key Words: Forecasting; Recession; Stagnation; ROC.

*Corresponding author. E-mail: `artur.tarassow@wiso.uni-hamburg.de`. Address for correspondence: Department of Socioeconomics, University of Hamburg, Welckerstr. 8, 20354 Hamburg, Germany.

1 Introduction

In recent years, and especially since the 2007-08 global financial crisis and the resulting worldwide economic recession, many countries have featured a prolonged period of low-growth coupled with stagnant high unemployment rates (Adler et al., 2017). Subsequently, the standard dichotomous classification of the business cycle in expansionary and contractionary phases seems to be somehow outdated for this new low-growth macroeconomic environment.

Following the work by Sichel (1994), a number of studies has already investigated the advantages of extending the characterization of the business cycle into a three-phase phenomenon. For instance, Krolzig and Toro (2001) study the interaction between U.S. real GDP and employment using a three-state Markov-switching (MS)-VAR model, where the three (unobservable) states are associated with recessions, "normal" and high growth periods. Similarly, Ferrara (2003) estimates a three-state MS-VAR for the US economy where the states are denoted as low, intermediate and high growth regimes, relative to the economy's trend growth rate, see also Nalewaik (2011) and Ho and Yetman (2012). By contrast, while Schreiber and Soldatenkova (2015) compute the probabilities for recessions, stagnations and expansions from a subset-VAR for the US economy, Proaño (2017) proposes a non-parametric dating algorithm for the identification of accelerations, "normal" or low economic growth periods and recessions, and predicts these phases for the German economy within an ordered probit model framework. Further, Candelon et al. (2013) investigate a four-regime ordered probit model allowing also for the differentiation between normal and severe recessions. Along the same lines Carstensen et al. (2017) use a three-state Markov-switching dynamic factor model for the prediction of ordinary and severe recessions.

An interesting advantage of a more differentiated characterization of the business cycle phenomenon highlighted by Nalewaik (2011) and Proaño (2017) is that specifying a low or stagnant growth regime leads to improvement in the forecasting performance concerning economic recessions, as the former may proceed or follow the latter. Therefore, a richer classification of the business cycle may have important advantages for policy-making in real time.

In this context, the contribution of this paper to the literature is the application of the three-regime business cycle characterization and prediction proposed by work by Proaño (2017) to two major industrialized economies which have experienced prolonged periods of low growth in recent times: the U.S. and Japan. Through a thorough analysis of the forecasting performance of the ordered probit model relative to the more standard binary probit model approach for different forecasting horizons we provide detailed insights on the value added of the ordered probit approach in economies where low growth phases are not seldom or negligible.

Our results can be summarized as follows: First, corroborating the previous findings by Nalewaik (2011) and Proaño (2017), we find for the U.S. and Japan that a three-phase ordered probit model has indeed a superior performance – in statistical terms – than an analogous binary probit model distinguishing only between recessions and accelerations for all considered forecast horizons. This is also true for the prediction of accelerations and normal or low growth periods (when compared with the corresponding binary probit models) for particular horizons in the U.S. and Japan, being this predictive superiority not statistically significant in the other cases. Further, binary probit models were not found to have a larger forecasting power. This holds irrespective of the forecast horizon or forecast evaluation samples analyzed.

The remainder of this paper is organized as follows. In Section 2 we discuss a non-parametric dating algorithm for the identification of high and low growth phases and recessions along the lines of Proaño (2017), as well as the ordered probit modeling approach used for the prediction of these three business cycle phases. In section 3 we discuss our empirical results concerning the U.S.A. and Japan, and compare the performance of the ordered probit model in forecasting economic recessions with that of standard binary probit models using various forecasting evaluation measures. Finally, we draw some concluding remarks from this study in section 4.

2 Methodology

2.1 The Non-Parametric Dating Algorithm

As previously mentioned, we employ in the following empirical study the non-parametric dating algorithm proposed by Proaño (2017) for the classification of the business cycle into economic accelerations, low growth phases and recessions. This algorithm consists of the following stages: First, the recessionary phases are identified according to the Harding and Pagan (2002) extension of the Bry-Boschan (1971) algorithm, whereafter potentially recessionary periods are those between a business cycle peak, defined as

$$\{y_{t-k} < y_t^p > y_{t+k}\ , \quad k = 1, ..., 5\} \tag{1}$$

where y_t is the two-month moving average of the business cycle reference series, and a business cycle trough, defined as

$$\{y_{t-k} > y_t^\tau < y_{t+k}\ , \quad k = 1, ..., 5\}. \tag{2}$$

As an additional censoring rule for the identification of recessionary periods, Harding and Pagan (2002) propose the use of the following measure of "severity" of an economic downturn j

$$S_j = 0.5 \times \text{Deepness}_j \times \text{Duration}_j \tag{3}$$

where

$$\text{Deepness}_j = |y_t^p - y_t^\tau|/y_t^p, \tag{4}$$

and *Duration* refers to the number of months between peak and trough of the economic downturn considered (see also Anas et al., 2008). A recessionary period is identified when $S_j > 0.025$, as there is no consensus on the reference minimum duration and deepness of recessions (Darné and Ferrara, 2009, p.5).

The second stage of the Proaño (2017) dating algorithm consists of identifying among the non-recessionary periods those which could be potentially considered as true economic accelerations or booms, in contrast to those which are periods of low or normal economic growth. For this purpose, the six-month moving average of the period-to-period growth rates of the reference series is calculated. An economic acceleration period is then identified if

a) the annualized *centered moving average* period growth rate in t exceeds a pre-determined and country/economy-specific value $\bar{g}_t^{min}$ which stands for the "normal" growth given population growth and technological progress, i.e.

$$\bar{g}_t = \frac{1}{6}\sum_{i=-3}^{3} g_{t-i}, \geq \bar{g}_t^{min} \qquad \text{with} \qquad g_t = 100 \cdot \left(\frac{Y_t}{Y_{t-1}} - 1\right) \tag{5}$$

b) the acceleration of Y_t is not lower than a given threshold, e.g.

$$\Delta g_t \geq \underline{\Delta g}_t^{min} \tag{6}$$

where Δ refers to the standard difference operator.

All periods which are not identified as acceleration or recessionary periods, are identified as periods of low economic growth.

2.2 Econometric Modeling

Analogously to Proaño (2017), we use an ordered probit model to estimate the three phases of the business cycle. For this purpose the following discrete variable is defined as follows:

$$c_t = \begin{cases} 0, & \text{if the economy goes through an economic recession,} \\ 1, & \text{if the economy goes through a stagnative growth phase, or} \\ 2, & \text{if the economy experiences an accelerative economic phase at time } t \end{cases} \tag{7}$$

with each of these outcomes being jointly determined by the non-parametric dating algorithm described in the previous section.

As discussed in Proaño (2017), the conditional probabilities of observing each outcome of c_t on the basis of a particular set of regressors $\mathbf{z}_t^i$ are given by

$$\Pr\left(c_{t+h} = 0|\mathbf{z}_t^i, \beta^i\right) = \Phi\left(0 - \mathbf{z}_t^{i\prime}\beta^i\right) \tag{8}$$

$$\Pr\left(c_{t+h} = 1|\mathbf{z}_t^i, \beta^i\right) = \Phi\left(1 - \mathbf{z}_t^{i\prime}\beta^i\right) - \Phi\left(0 - \mathbf{z}_t^{i\prime}\beta^i\right) \tag{9}$$

$$\Pr\left(c_{t+h} = 2|\mathbf{z}_t^i, \beta^i\right) = 1 - \Phi\left(2 - \mathbf{z}_t^{i\prime}\beta^i\right) \tag{10}$$

with $\Phi(\cdot)$ being the cumulative normal distribution function employed in ordered probit models. As in Proaño (2017), the set of regressors $\mathbf{z}_t^i$ is restricted by the real-time data availability of the included variables and is parsimoniously determined through an automatic General-to-Specific (G2S) indicator selection procedure based on the statistical significance of the individual coefficients which reduces the dimensionality of the regression model.[1]

In the following forecasting exercise the ordered probit model is re-estimated in each period using the most recent data (in a rolling window fashion) up to that period (under real-time conditions), so that all h-step ahead forecasts are pseudo out-of-sample, i.e. they are based only on values of the series up to the date on which the forecast is made. The h-step ahead forecast will be computed by means of the direct forecast method, what implies the estimation of a separate regression for each forecast horizon h.

In order to evaluate the forecasting performance of the ordered probit model concerning particular business cycle phases, the three business cycle phases are estimated separately by more standard dynamic probit regressions as in Proaño (2017). Accordingly, while the binary recession series b_t is set such that

$$b_t = \begin{cases} 0, & \text{if the economy experiences an expansion or a low growth phase at time } t, \\ 1, & \text{if the economy goes through a recessionary phase at time } t \end{cases} \tag{11}$$

the binary acceleration variable a_t is defined as

$$a_t = \begin{cases} 0, & \text{if the economy goes through a recessionary or low growth phase at time } t, \\ 1, & \text{if the economy experiences an accelerative economic phase at time } t. \end{cases} \tag{12}$$

and the low growth variable l_t is defined as

$$l_t = \begin{cases} 0, & \text{if the economy goes through a recessionary or an acceleration phase at time } t, \\ 1, & \text{if the economy experiences a low growth phase at time } t. \end{cases} \tag{13}$$

These three binary series b_t and a_t and l_t are then estimated separately using exactly the same set of explanatory variables as potential regressors, and indicator selection procedure.

[1] More specifically, the following algorithm is applied:

(i) Specify the model using the complete set of potential regressors.

(ii) Check whether any of the variables has a p-value larger than the significance level α.

(iii) Omit the variable with the highest p-value, p^*, if $p^* > \alpha$ and re-estimate the reduced model specification. If the p-value of none of the variables exceeds α, stop the algorithm. Otherwise repeat steps (ii) and (iii).

The final model consists only of the variables for which the corresponding significance level of the final model fulfills the condition $p^* \leq \alpha$. In our work we set $\alpha = 0.05$. Alternatively, we performed the complete analysis using a specific-to-general (S2G) lag selection procedure which delivered quite robust results. All `Gretl` (Cottrell and Lucchetti, 2017) codes required for the replication of our results are available upon request.

2.3 Forecast Evaluation Criteria

We evaluate the forecasting performance of the ordered and the binary probit models at different forecast horizons using a variety of evaluation measures. Two standard overall measures of probability forecast quality which have been recently applied e.g. by Lahiri and Wang (2013) and Döpke et al. (2015) are Brier's *Quadratic Probability Score* (QPS) and the *Log Probability Score* (LPS) over the spectrum of forecasts of interest (Diebold and Rudebusch, 1989). The QPS is a measure of the mean squared error comparing the predicted probability of an event with an indicator of the event, and is defined as

$$QPS = T^{-1} \sum_{t=1}^{T} \left(\widehat{Y}_t - Y_t\right)^2 \tag{14}$$

where $\widehat{Y}_t$ refers to the *ex-ante* probability of an event at time t and Y_t is a binary variable taking on the value unity at period t when the event of interest occurs, otherwise zero. T is the total number of forecasts available. The QPS takes a score of 0 in case of perfect accuracy and 1 *vice versa*.

In contrast, the LPS is defined as

$$LPS = -T^{-1} \sum_{t=1}^{T} \left[(1 - Y_t)\ln(1 - \widehat{Y}_t) + Y_t \ln(\widehat{Y}_t)\right] \tag{15}$$

and ranges from 0 to ∞, with the value of 0 meaning perfect forecasting accuracy. Note that the LPS penalizes large errors more heavily than QPS.

An alternative approach to assess probability forecast performance is to focus on the hit rate and the false alarm rate (Lahiri and Wang, 2013, 175). Consider the standard contingency table (see Table 1) for binary outcomes of the realizations Y and its associated probability forecast $\hat{Y}$ where $Y = 1$ represents the actual occurrence of an event (otherwise 0) and $\hat{Y} = 1$ the correct prediction of the occurrence of this event (otherwise 0)

Table 1: Contingency table for binary outcomes

		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

where *TN*, *FN*, *FP* and *TP* denote the number of true negative, false negative, false positive and true positive. The total number of the occurrence of actual events is $P = TP + FN$ while the total number of no events is given by $N = TN + FP$. The *receiver-operating curve* (ROC) plots the true positive rate (TP/P) against the false positive rate (FP/N) for each value of $\hat{Y}$ showing that both rates are functions of the cut-off ω^*, also called the binary event classifier. For a skillful forecast both the true positive rate, and the *specificity*, one minus the false positive rate, are expected to be high.

We compute two summarizing statistics of the ROC curve: (i) the AUROC, and (ii) the Youden index. More specifically, the AUROC is the area under the ROC curve, and measures the overall accuracy of the forecast model. A perfect classifier has an AUROC of 1 while a model with an AUROC of 0.5 predicts no better than a coin flip. In the following, we will test the null hypothesis of no difference between the AUROC statistics of two competing models, $H_0 : AUROC_i - AUROC_j = 0,\ i \neq j$, see DeLong et al. (1988).

However, the AUROC does not provide information on the optimal cut-off ω^*. By contrast, the Youden index is defined as the distance between the hit rate and the false alarm rate as measured by the maximal vertical gap between the diagonal and the ROC curve. This distance is just maximized at a cut-off value which maximizes the fraction of correctly predicted binary outcomes which are functions of ω^*. Thus, the Youden index not only provides a numerical summary of the ROC curve, but it also measures the *local* skill of a specific cut-off. This statistics is computed as

$$\text{Youden} = \frac{TP}{TP + FP} + \frac{TN}{TN + FP} - 1\,. \tag{16}$$

A Youden index of zero corresponds to a random classifier while a Youden of 1 refers to a perfect classifier.[2]

3 Empirical Analysis

3.1 Data Description

For the U.S. we use the monthly FRED-MD dataset as provided by McCracken and Ng (2015) which comprises monthly data between 1977m1 and 2016m9 (480 observations). The Japanese macroeconomic and financial data are obtained from the Bank of Japan and Datastream. The vector of exogenous regressors for the U.S. includes growth of real manufacturing new orders for durable goods, growth of industrial production, S&P500 real stock market returns, the 10-year Treasury rate minus the 3-month Treasury bill rate, the 3-month commercial paper rate minus the Federal funds rate, growth of real oil prices, and the consumer sentiment index. Japanese monthly data cover the period between 1989m1 and 2015m12 (324 observations). We consider growth of industrial production, growth of real domestic machinery orders, growth of foreign machinery orders, growth of real retail trade, Nikkei real stock market returns, the difference between the 10-year government bond yield and the 3-month money market rate, growth of real oil prices, the business sentiment index, and growth of unfilled job vacancies. For details on the data and eventual transformations we refer to the Data Appendix.

[2]For computation of the ROC curves and related statistics we use the `Gretl` package `roc` (vers. 1.02) provided by Peter M. Summers.

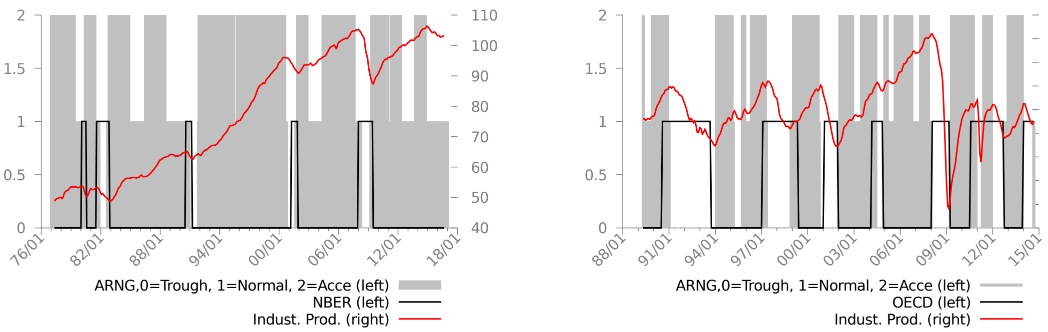
For the parameters $\bar{g}_t^{min}$ and $\Delta \underline{g}_t^{min}$ of the dating algorithm proposed by Proaño (2017) we use the following values for the U.S. and Japan, respectively, as described in Table 2.

Table 2: Overview of parameter settings for the construction of the ARNG indicator.

Parameter	USA	Japan
$\bar{g}_t^{min}$	0.003	0.002
$\Delta \underline{g}_t^{min}$	-0.015	-0.015

The value $\bar{g}_t^{min} = 0.003$ (implying an annual growth of potential output of about $(1+0.003)^{12} - 1 \approx 3.6\%$) has its roots in the classical Okun's law relationship. For Japan we set this parameter a bit lower to $\bar{g}_t^{min} = 0.002$ as a consequence of lower productivity growth during the last decades. For comparison, a value of $\bar{g}_t^{min} = 0.0025$ for Germany was chosen in Proaño (2017). By contrast, the value for $\Delta \underline{g}_t^{min}$ is data-driven and is chosen as to avoid a too volatile switching between the acceleration and the low growth periods (in Proaño, 2017 the value of 0.001 was chosen for Germany).

Figure 1 depicts the underlying industrial production series and the corresponding business cycle classification using the ARNG indicator for both the U.S. and Japan, respectively. For the U.S. we additionally plot the NBER recession dates (which are based on real GDP growth instead), while for Japan we add estimated OECD recession dates (also based on real GDP growth). Overall, in both countries the ARNG algorithm delivers a plausible chronology of the business cycle. On the one hand, the recessionary periods are in line with the NBER and OECD recession dating for the U.S. and Japan, respectively. Further, the differentiation between true economic accelerations and low growth periods seems particularly meaningful in the U.S. in some periods after the 2007 crisis, as well as in Japan in numerous periods over the whole sample.



(a) USA

(b) Japan

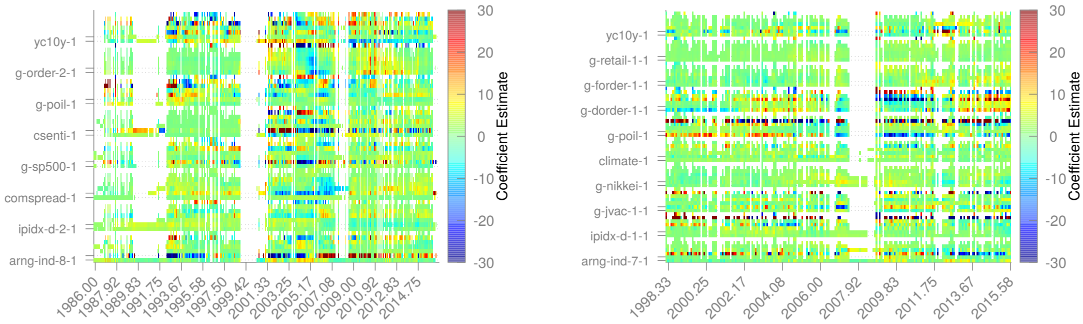
Figure 1: Dating of economic phases as defined in Table 2. ARNG dating based on U.S. and Japanese industrial production index (two-month moving average). For the U.S. the NBER dating is based on growth of real GDP, and for Japan the recession dates are based on OECD estimates using growth of real GDP as the reference series.

3.2 Estimation Results

In this section we discuss the out-of-sample probability forecasts of both the dynamic ordered probit model as well as dynamic binary probit models using the ARNG indicator.

For the U.S., the complete sample ranges between 1979m1 and 2016m9, and the initial training set uses 84 observations from 1979m1 to 1985m12 to determine the optimal set of features (regressors) before computing the h-step-ahead direct forecasts. Next, the beginning and end of the training set are extended by one additional observation such it ranges from 1979m2 to 1986m1. The last training set is determined before computing the h-multi-step direct forecasts. The number of forecast sequences is 369. For Japan the sample ranges from 1990m4 to 2015m9, and the initial training set uses 84 observations from 1991m6 to 1998m5 yielding 209 forecasting sequences.[3]

As previously mentioned, in each of the following periods the ordered and the binary probit models are parsimoniously re-specified through an automatic General-to-Specific (G2S) indicator selection procedure starting with a maximum of seven lags of each variable. This implies that the set of regressors in the predicting models may vary significantly over time, as well as the recurrently newly estimated individual coefficients. Figure 2 illustrates the normalized (by their respective full-sample median value) point coefficient estimates from the ordered probit model using a rolling-window estimation method.



(a) U.S. (b) Japan

Figure 2: Rolling-window (relative) coefficient estimates after automatic indicator-selection from the dynamic ordered probit regression based on the ARNG indicator. Reported results are based on the 1-month-ahead direct forecasting regressions.

As it can be clearly observed in Figure 2, there is a significant variability concerning the value of the estimated parameters and of the actual variables over the different vintages in both the United States and Japan, what highlights the meaningfulness of re-estimating the regression model in every period when new information becomes available.

Interestingly, for the U.S. we observe that during the three low-growth/recession episodes

[3] For both the U.S. and Japan the results are robust against the use of a rolling-window of width 96 monthly observations.

namely between the late 1980 and beginning of the 1990s, between 1999 and 2000, and in 2008 only a sub-set of relevant indicators were selected. While lags of the ARNG indicator and the spread between the 10-year Treasury rate minus the 3-month Treasury bill rate are frequently selected throughout the whole sample, we observe the relevance of growth of industrial production and consumer sentiments in the late 1980s while the commercial paper spread, growth of real stock market returns and also consumer sentiments are frequently selected during the Great Financial Crisis (GFC) period in 2008. For Japan, we find that for most periods all features (but not all lags) are selected by the G2S algorithm. However, during the GFC period, we find that lags of the ARNG indicator, growth of industrial production, real stock market returns and business sentiments are frequently among the selected features.

Figure 3 illustrates the out-of-sample one-month ahead probability forecasts for the period between 2007m1 and 2010m12 for the three business cycle phases determined by the dating algorithm proposed by Proaño (2017). Panel A shows how timely the ordered probit model is able to identify both the beginning as well as end of the recession period in the U.S. and Japan.

For instance, even though both models miss the exact starting date of the recession in 2008, the ordered probit model signals 1-2 months earlier the recession compared to the binary probit model for both the U.S. and Japan. Also the binary probit model wrongly signals with high probability a recession in autumn 2009 and May/June 2010 in the U.S., and in May 2009 and January 2010 in Japan. For the U.S. we also observe that the binary probit model wrongly predicts low growth periods in May 2008 as well as January 2010, and misses the low growth period in summer 2009. Lastly, the ordered probit model identifies the starting of the acceleration phase in autumn 2009 1-2 months earlier in the U.S., and also provides more stable probability forecasts during this acceleration period for both Japan and the U.S.

3.2.1 Recession Probability Forecast Evaluation

We evaluate first the recession prediction performance of the binary as well as the ordered probit models. Figure 4 illustrates both the LPS and QPS measures for both models for (i) the complete forecasting sample, and (ii) for a forecasting sample starting in 2007m1 to focus on the forecasting performance of both models since the GFC period.

For the whole forecasting sample we find for both the U.S. and Japan that the ordered probit model yields more accurate forecasts compared to the binary probit model according to both the QPS and LPS criteria. More specifically, for the U.S. the LPS measure of the binary probit model is about three times higher compared to the ordered probit model for the first three horizons and even 5 times higher for the 6-month horizon. For Japan, we report at the 1-month horizon a LPS of 0.19 for the ordered probit model compared to a LPS of 0.49 for the binary probit model. This performance gap remains for all forecast horizons. For the QPS measure qualitatively similar observations arise. For the forecasting subsample starting

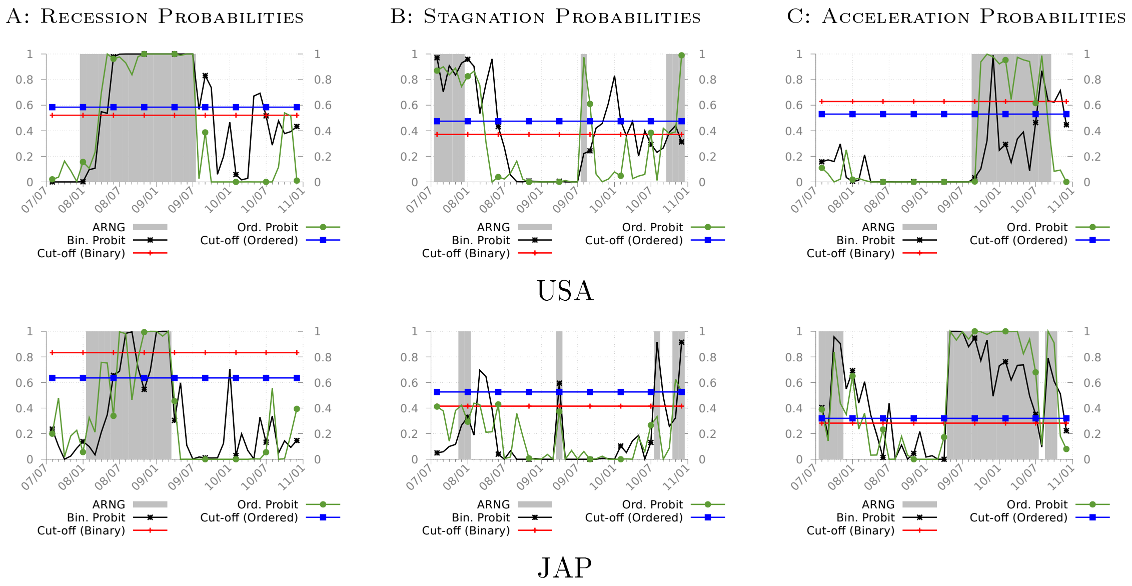


Figure 3: Out-of-sample 1-month ahead state-probabilities during the Great Financial Crisis. The gray shaded area displays the respective business cycle state as identified by the ARNG dating algorithm. The optimal cut-off value ($\omega*$) is the value which maximizes the fraction of correctly predicted outcomes of the dependent variable.

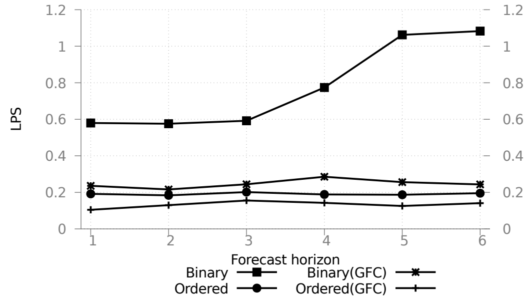
in 2007m1, the ordered probit model still outperforms the binary probit alternative for both forecast accuracy criteria and each horizon. However, at least for the U.S. we observe that the advantage in terms of forecast accuracy of the ordered probit model is somewhat lower now.

Apart from the QPS and LPS measure, we evaluate the recession probability forecasts for both countries using the AUROC statistics and the Youden index. The results for the full sample and the subsample since 2007m1 are reported in Figure 5.

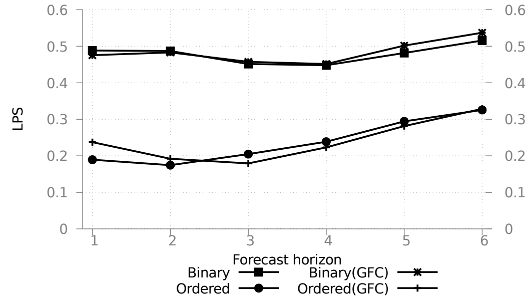
The results indicate that the ordered probit model is superior to the binary probit model for both the U.S. and Japan. Based on the complete forecasting sample for the U.S., the ordered probit yields an AUROC ranging, dependent on the forecast horizon, between 0.85 and 0.89. In the binary probit case, this statistics is about 0.84 for the first three horizons but falls substantially to 0.65 at the 6-month horizon which confirms its rather poor forecasting skills. These findings are also confirmed by the Youden index. Interestingly, when considering only forecasts since 2007, we observe that both models yield slightly higher AUROC and Youden statistics throughout all forecast horizons. However, still the ordered probit model dominates.

For Japan, based on all forecasts available, we observe that the ordered probit model yields at each horizon an AUROC statistics of about 0.96 while the binary probit model reaches values ranging between 0.84 and 0.86. Similarly, we find that the Youden index of the ordered probit model outperforms the binary probit alternative at each horizon while that performance gap even increases the longer the forecast horizon gets. Interestingly, for the sample since 2007m1 we observe a slight improvement in forecasting skills for the binary probit model according

Panel A: Log Probability Score

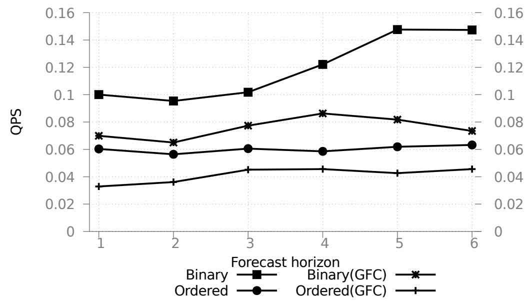


(a) USA

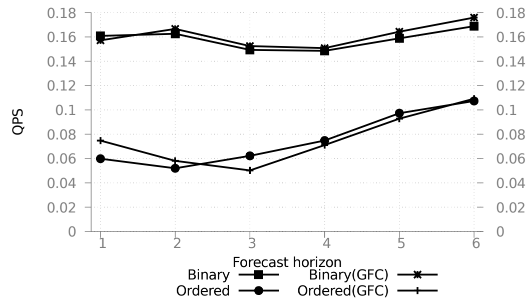


(b) JAP

Panel B: Quadratic Probability Score



(c) USA



(d) JAP

Figure 4: *Recession* probability forecasts evaluation criteria. 'Binary' and 'Ordered' refer to the forecasts evaluation criteria based on the binary probit and ordered probit models using all observations available, respectively. The abbreviation 'GFC' denotes the respective statistics only based on observations since 2007m1.

to both the AUROC and the Youden statistics while the Youden index of the ordered probit model deteriorates. However, the ordered probit model still dominates throughout all forecast horizons.

Table 3 reports the results of the test on equal AUROC statistics between the binary probit model and the ordered probit model for each forecast horizon. In the following, we set a significance level of 5% for rejecting the null hypothesis.

For the U.S. we find evidence that, using all forecasts available, the ordered probit model significantly outperforms the binary probit alternative at all forecast horizons as the null hypotheses can be rejected at the 1% level. Considering only the period since 2007m1, we can reject the null at least for forecast horizons longer than three months. These results underline that the ordered probit case yields significantly better probability forecast results compared to the binary probit case in the U.S. Very similar results are obtained for Japan, as, independent of the chosen sample, the null can be rejected for each forecast horizon.

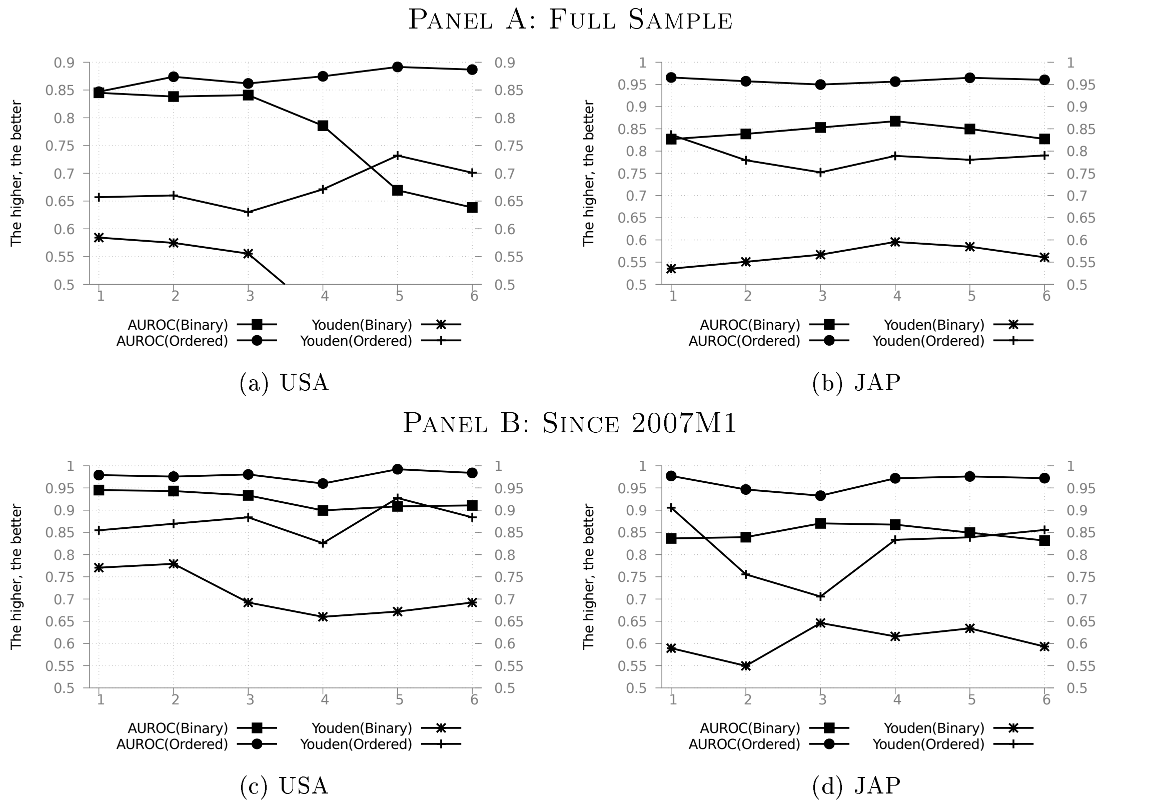


Figure 5: AUROC statistics and Youden index of out-of-sample *recession* probability forecasts based on the ARNG indicator.

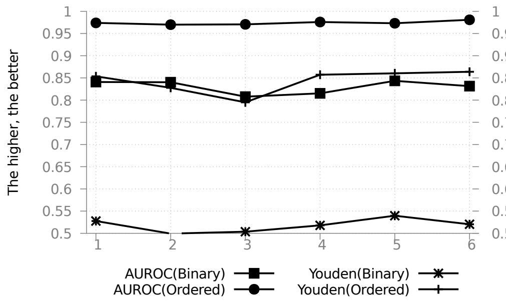
Table 3: Test on *Area Under the ROC*-equality of out-of sample *recession* probability forecasts between the ordered-probit and the binary-probit model by horizon.

	Test (USA)	p-value (USA)	Test (JAP)	p-value (JAP)
(A) Full sample				
h=1	17.258	0.000	23.274	0.000
h=2	17.217	0.000	23.262	0.000
h=3	14.873	0.000	17.873	0.000
h=4	31.411	0.000	17.206	0.000
h=5	76.935	0.000	6.983	0.008
h=6	81.077	0.000	8.238	0.004
(B) Since 2007m1 only				
h=1	2.822	0.093	8.492	0.004
h=2	1.926	0.165	11.270	0.001
h=3	2.588	0.108	9.578	0.002
h=4	4.391	0.036	9.896	0.002
h=5	6.430	0.011	5.509	0.019
h=6	4.909	0.027	5.958	0.015

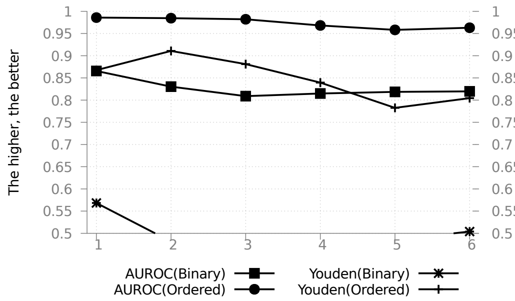
3.2.2 Acceleration Probability Forecast Evaluation

Analogously to Figure 5, Figure 6 shows the AUROC and Youden index values for the prediction of acceleration periods through the ordered probit and a corresponding binary probit model. Table 6 in the Appendix summarizes the corresponding test on equal AUROC statistics between the binary probit model and the ordered probit model for each horizon.

Panel A: Full Sample

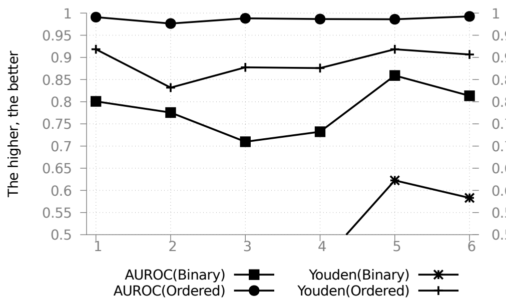


(a) USA

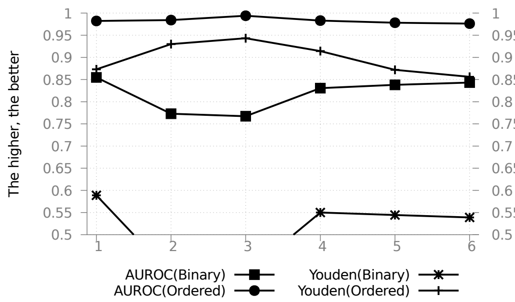


(b) JAP

Panel B: Since 2007M1



(c) USA



(d) JAP

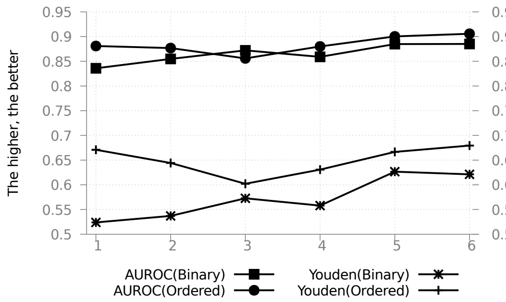
Figure 6: AUROC statistics and Youden index of out-of-sample *acceleration* probability forecasts based on the ARNG indicator.

For the United States, our results indicate that the ordered probit seems to have a superior forecasting performance than a binary probit model over all horizons and in both forecasting samples applied on an acceleration binary series concerning the AUROC and Youden index values. However, the forecasting performance of both prediction models is not statistically different at all forecasting horizons when the whole forecasting sample is considered. By contrast, the opposite holds when the post-crisis forecasting subsample (when the low growth phases are relatively more relevant) is considered. This confirms the superiority of the ordered probit model for predicting acceleration phases since 2007. Similarly, for Japan, where low growth phases are more frequent and thus a differentiated treatment is more meaningful, the results of the test for forecasting equivalency indicate that the forecasts of the ordered probit model are statistically superior than those of a binary probit model. This holds for both samples tested.

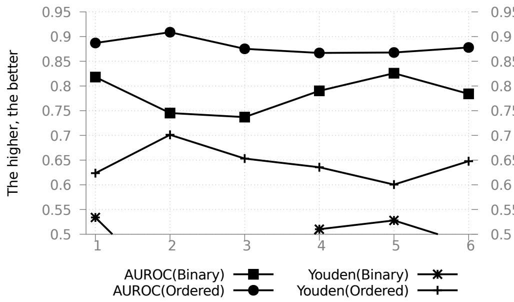
3.2.3 Low Growth Probability Forecast Evaluation

Analogously, Figure 7 illustrates the AUROC and Youden index values for the prediction of low growth periods by the ordered probit and a binary probit model for that particular regime. The corresponding test results for forecasting equivalency are summarized in Table 7 in the Appendix.

Panel A: Full Sample

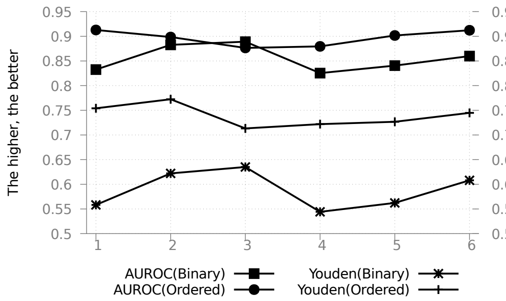


(a) USA

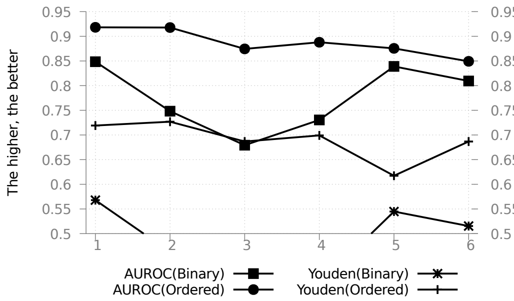


(b) JAP

Panel B: Since 2007M1



(c) USA



(d) JAP

Figure 7: AUROC statistics and Youden index of out-of-sample *low growth* probability forecasts based on the ARNG indicator.

As in the previous case, the AUROC and the Youden index values for the U.S. indicate that the forecasting performance of the ordered probit model is superior to that of the binary probit for nearly all forecast horizon in both forecast evaluation periods. However, the test results in Table 7 indicate that only for the one-month ahead forecast horizon in the post-crisis evaluation sample the ordered probit model's performance is statistically different, in terms of the AUROC statistics, than that of the binary model (rejection at the 10% level). We obtain a similar picture for Japan, where the ordered probit model's forecasting performance is only statistically superior than that of the binary model at certain forecast horizons. Nevertheless, in terms of the Youden index the ordered probit model dominates throughout all horizons.

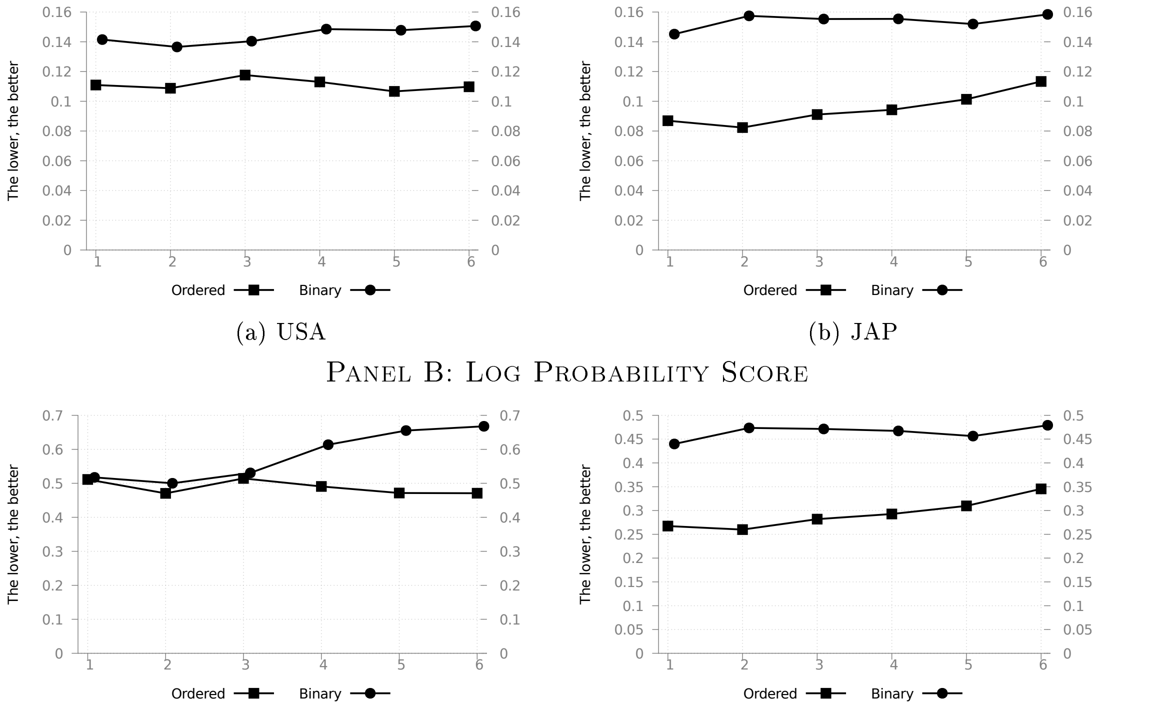
3.2.4 Overall probability forecast evaluation

In the previous subsection we evaluated the recession, acceleration and low growth probability performance separately for both the binary and ordered probit model, respectively. In this section, we go a step further and assess the skills of each model type to forecast jointly recession periods, low growth periods as well as acceleration events.

To do so, we compute both the QPS and LPS statistics for the recession probability binary probit model (as already discussed before), the low growth probability binary probit model and the acceleration probability binary probit model. Using these three QPS and LPS statistics, we compute the mean LPS and QPS statistics across all three states. The same is done for the ordered probit model. Hence, we obtain a joint forecast accuracy measure for the binary and the ordered probit model, respectively. Figure 8 reports the joint measures using all forecasts available.

Panel A: Quadratic Probability Score

Panel B: Log Probability Score



(a) USA (b) JAP

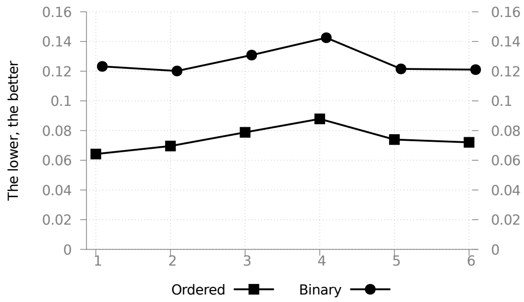
(c) USA (d) JAP

Figure 8: Joint probability forecast evaluation of *recessionary*, *low growth* and *acceleration* periods using all forecasts available.

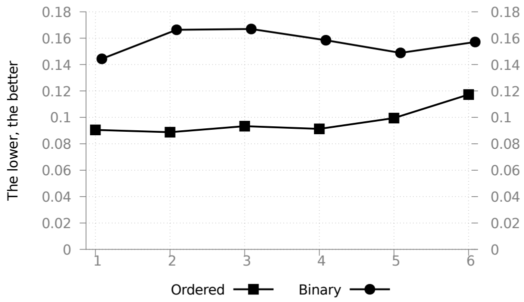
For the U.S. we see that the ordered probit case yields an QPS of about 0.11 compared to an average of 0.15 for the binary probit case across all six forecast horizons. Interestingly, in terms of the LPS measure both models perform similarly for the first three forecast horizons before the ordered probit model dominates at longer horizons. This implies that the binary probit model suffers from some large forecast errors at longer horizons which the LPS penalizes more heavily compared to the QPS. In the Japanese case we observe that the ordered probit model clearly outperforms the binary probit model at all horizons according to both criteria. The performance advantage of the ordered probit model is at least about 40% to 50% for both the QPS and the LPS measure, respectively.

In Figure 9 we illustrate the same statistics for the forecasting sample since 2007m1. While the results remain unchanged for the Japanese case, we find even stronger hints for the U.S. that the ordered probit model dominates the binary probit model stressing the good joint forecast accuracy performance of the ordered binary model since 2007.

Panel A: Quadratic Probability Score

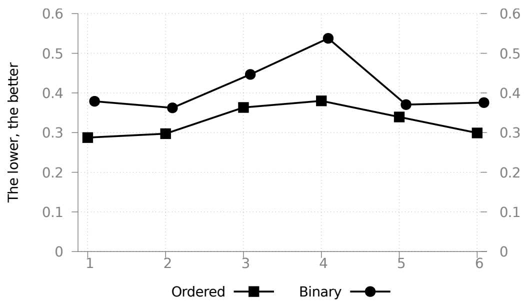


(a) USA

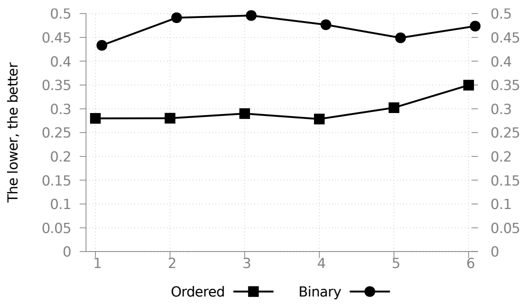


(b) JAP

Panel B: Log Probability Score



(c) USA



(d) JAP

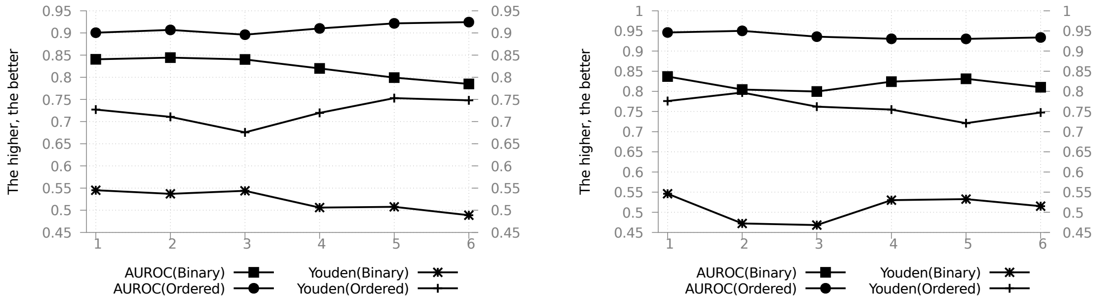
Figure 9: Joint probability forecast evaluation of *recessionary*, *low growth* and *acceleration* periods since 2007m1.

Next, we also compute a joint measure for the AUROC and Youden statistics in the similar vein as we have done before for the LPS and QPS measures. The results are depicted in Figure 10.

For the U.S., using all forecasts available, we find that the ordered probit model yields higher joint AUROC as well as joint Youden statistics throughout all forecast horizons confirming the previous findings. For instance, at the 1-(6-)month horizon, the ordered probit model yields an AUROC of 0.9 (0.94) compared to an AUROC of 0.84 (0.79) obtained by the binary probit model. Interestingly, when considering forecasts since 2007m1, we observe that both models yield slightly higher joint AUROC statistics even though their relative performance gap remains stable throughout all horizons. However, the binary probit model shows substantial weaknesses in classification skills according to the joint Youden index for forecasts longer than 2-months ahead.

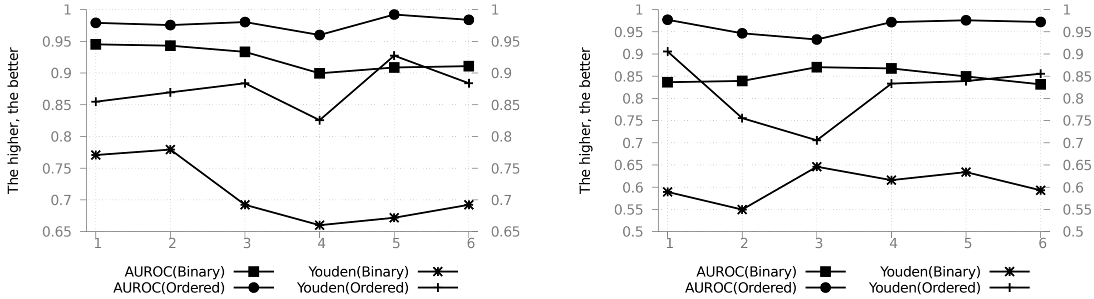
Using all available forecasts for Japan, we can confirm the dominance of the ordered probit model at each forecast horizon. For instance, while the ordered probit models yields a joint AUROC statistics of about 0.95 throughout all horizons, the binary probit yields only a joint AUROC of 0.84. The performance gap is even larger for the joint Youden index (0.75 against 0.53) irrespective of horizons considered. The results remain unchanged when considering only forecasts since 2007. In total, the results indicate for both the U.S. and Japan that the ordered probit model framework works remarkably well relative to the binary probit case in terms of probability forecast accuracy as well as classification.

Panel A: Full Sample



(a) USA (b) JAP

Panel B: Since 2007M1



(c) USA (d) JAP

Figure 10: Joint probability forecast evaluation by means of AUROC and Youden statistics.

4 Concluding Remarks

How many are enough? So far, the great majority literature on business cycle forecasting has used the dichotomous characterization of the business cycle consisting in recessionary and expansionary phases stemming from the seminal work by Burns and Mitchell (1946). However, many countries around the world have experienced pronounced periods of stagnant or low economic growth coupled with high and persistent unemployment rates and slow technological progress which can barely be considered as expansionary phases in recent times. Against this background, new research has explored more differentiated classifications of the business cycle phenomenon.

In this paper we applied the three-phase business cycle classification by Proaño (2017) – accelerations, low growth periods and recessions – and used his proposed ordered-probit approach to forecast these economic phases for two major industrialized economies: the U.S. and Japan. Using state-of-the-art forecast evaluation measures we assessed the forecasting power of the ordered probit model at various forecast horizons. Interestingly, and corroborating the results by Nalewaik (2011) for the U.S. and Proaño (2017) concerning Germany, a more differentiated characterization of the business cycle phenomenon in the U.S. and Japan and the subsequent prediction of multiple business cycle phases in a consistent framework – where the three estimated state probabilities add up to one in each period – turned to be superior in many cases to the performance of binary models aimed at forecasting each of the three phases individually.

Summing up, our results suggest that the approach pursued in this paper is a promising direction for future research aimed at a better characterization of the business cycle phenomenon.

References

Adler, G., Duval, R. A., Furceri, D., Celik, S. K., Koloskova, K., and Poplawski-Ribeiro, M. (2017). Gone with the Headwinds; Global Productivity. IMF Staff Discussion Notes 17/04, International Monetary Fund.

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

Bry, G. and Boschan, C. (1971). *Cyclical Analysis of Time Series: Selected Procedures and Computer Programs.* NBER Books. National Bureau of Economic Research, Inc.

Burns, A. F. and Mitchell, W. C. (1946). *Measuring Business Cycles.* NBER.

Candelon, B., Metiu, N., and Straetmans, S. (2013). Disentangling economic recessions and depressions. Discussion Paper 43/2013, Deutsche Bundesbank, Frankfurt a.M.

Carstensen, K., Heinrich, M., Reif, M., and Wolters, M. H. (2017). Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model. An application to the German business cycle. Working Paper 6457, CESifo.

Cottrell, A. and Lucchetti, R. (2017). *Gretl User's Guide.*

Darné, O. and Ferrara, L. (2009). Identification of slowdowns and accelerations for the Euro Area economy. Working Paper 42/2009, CEPR/EABCN.

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44:837–845.

Diebold, F. X. and Rudebusch, G. D. (1989). Scoring the Leading Indicators. *The Journal of Business*, 62(3):369–91.

Döpke, J., Fritsche, U., and Pierdzioch, C. (2015). Predicting recessions in Germany with boosted regression trees. Macroeconomics and Finance Series 201505, Hamburg University, Department Wirtschaft und Politik.

Ferrara, L. (2003). A three-regime real-time indicator for the US economy. *Economics Letters*, 81:373–378.

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

Ho, W.-Y. A. and Yetman, J. (2012). Does US GDP stall? Working Paper 387, BIS, Basel.

Krolzig, H.-M. and Toro, J. (2001). A new approach to the analysis of business cycle transitions in a model of output and employment. Department of Economics Discussion Paper 59, Oxford University, Oxford.

Lahiri, K. and Wang, J. G. (2013). Evaluating probability forecasts for GDP declines using alternative methodologies. *International Journal of Forecasting*, 29(1):175–190.

McCracken, M. W. and Ng, S. (2015). FRED-MD: A Monthly Database for Macroeconomic Research. Working Papers 2015-12, Federal Reserve Bank of St. Louis.

Nalewaik, J. J. (2011). Forecasting recessions using stall speeds. Finance and Economics Discussion Series 2011-24, Federal Reserve Board, Washington, D.C.

Proaño, C. R. (2017). Detecting and predicting economic accelerations, recessions, and normal growth periods in real-time. *Journal of Forecasting*, (1):26–42.

Schreiber, S. and Soldatenkova, N. (2015). Anticipating business-cycle turning points in real time using density forecasts from a VAR. *Journal of Macroeconomics*, 47:166–187.

Sichel, D. E. (1994). Inventories and the three phases of the business cycle. *Journal of Business & Economic Statistics*, 12(3):269–277.

A Appendix

Table 4: U.S. macroeconomic and financial indicators

Series	FRED-MD + Transformation	Description	Vintage Data	Publication Lag
IPIDX-D	$100 \times \left(\frac{INDPRO_t}{INDPRO_{t-1}} - 1\right)$	Growth of industrial production	yes	2 months
G-ORDER	$100 \times \left(\frac{\frac{AMDMNOx_t}{CPIAUCSL_t}}{\frac{AMDMNOx_{t-1}}{CPIAUCSL_{t-1}}} - 1\right)$	Growth of real manufact. new orders of durable goods	yes	2 months
G-SP500	$100 \times \left(\frac{\frac{SP500_t}{CPIAUCSL_t}}{\frac{SP500_{t-1}}{CPIAUCSL_{t-1}}} - 1\right)$	S&P 500 real stock market return	no	0 months
YC10Y	$GS10 - TB3MS$	10-year Treasury rate minus 3-month Treasury bill	no	0 months
COMSPREAD	COMPAPFFx	3-Month Commercial Paper Minus Federal funds rate	no	0 months
G-POIL	$100 \times \left(\frac{\frac{OILPRICE_t}{CPIAUCSL_t}}{\frac{OILPRICE_{t-1}}{CPIAUCSL_{t-1}}} - 1\right)$	Real oil price growth	no	0 months
CSENTI	UMCSENTx	Consumer sentiment index	no	0 months

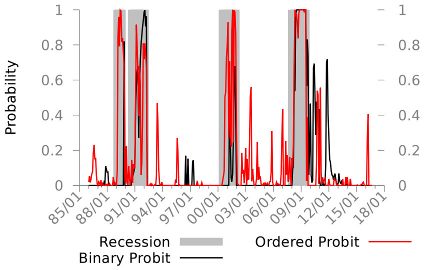
Sources: All series are obtained from the FRED-MD database McCracken and Ng (2015). All abbreviations used are in line with the FRED-MD database.

Table 5: Japanese macroeconomic and financial indicators

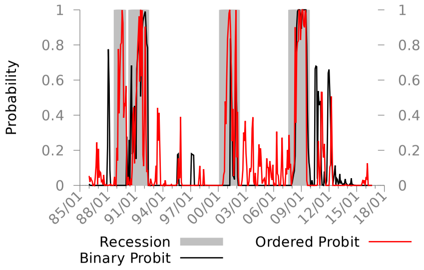
Series	Transformation	Description	Vintage Data	Publication Lag
IPIDX-D	$100 \times \left(\frac{INDPRO_t}{INDPRO_{t-1}} - 1\right)$	Growth of industrial production	yes	1 months
G-DORDER	$100 \times \left(\frac{\frac{order_t}{P_t}}{\frac{order_{t-1}}{P_{t-1}}} - 1\right)$	Growth of real domestic machinery orders (deflated by producer price index)	yes	1 months
G-FORDER	$100 \times \left(\frac{\frac{forder_t}{P_t}}{\frac{forder_{t-1}}{P_{t-1}}} - 1\right)$	Growth of real foreign machinery orders (deflated by producer price index)	yes	1 months
G-RETAIL	$100 \times \left(\frac{\frac{retail_t}{P_t}}{\frac{retail_{t-1}}{P_{t-1}}} - 1\right)$	Growth of real retail trade (volume) (deflated by producer price index)	yes	1 months
G-NIKKEI	$100 \times \left(\frac{\frac{nikkei_t}{P_t}}{\frac{nikkei_{t-1}}{P_{t-1}}} - 1\right)$	Nikkei real stock market return (deflated by producer price index)	no	0 months
YC10Y		10-year govern. bond yield minus 3-month money market rate	no	0 months
CLIMATE		Business sentiment index	no	0 months
POIL_DLN	$100 \times \left(\frac{\frac{OILPRICE_t}{CPIAUCSL_t}}{\frac{OILPRICE_{t-1}}{CPIAUCSL_{t-1}}} - 1\right)$	Real oil price growth	no	0 months
G-JVAC		Growth of unfilled vacancies	yes	1 months

Sources: All series are obtained either from the Bank of Japan or from Datastream

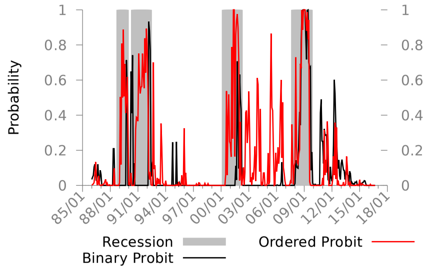
Panel A: U.S.



(a) 1-month forecast

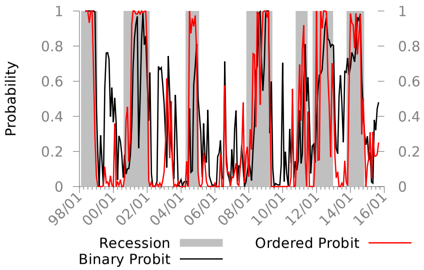


(b) 3-month forecast

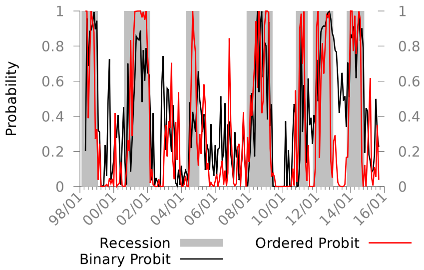


(c) 6-month forecast

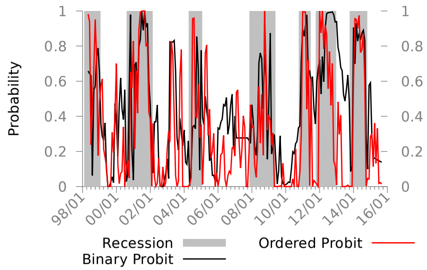
Panel B: JAP



(d) 1-month forecast



(e) 3-month forecast

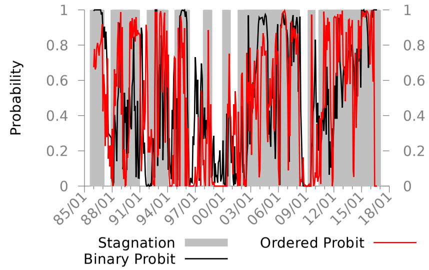


(f) 6-month forecast

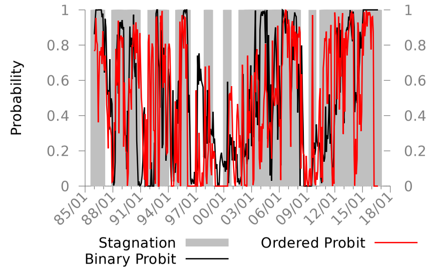
Notes: These are the out-of-sample recession probability forecasts using a rolling-window of width 84 monthly observations. The binary recession series is based on the ARNG indicator.

Figure 11: U.S. and JAP out-of-sample *recession* probabilities from both the dynamic binary and dynamic ordered probit models for different forecast horizons.

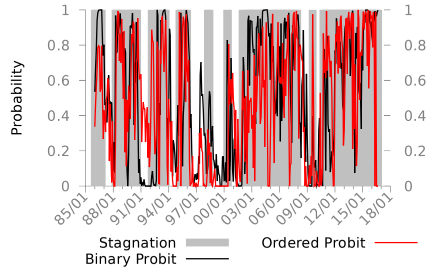
Panel A: U.S.



(a) 1-month forecast

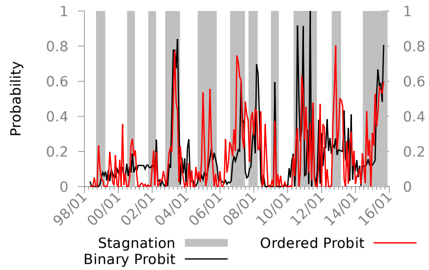


(b) 3-month forecast

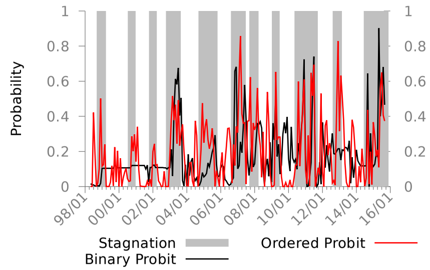


(c) 6-month forecast

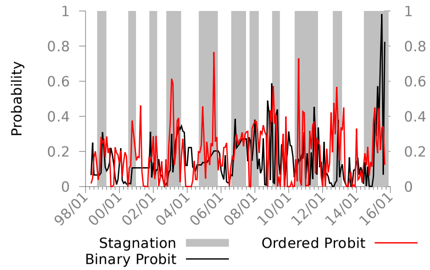
Panel B: JAP



(d) 1-month forecast



(e) 3-month forecast

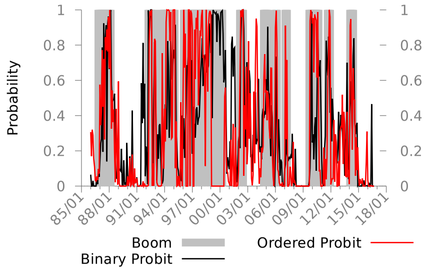


(f) 6-month forecast

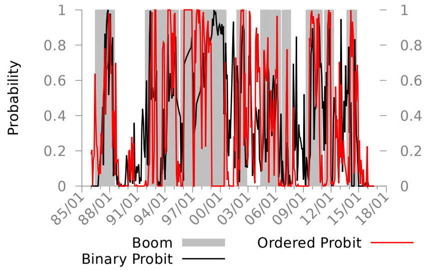
Notes: These are the out-of-sample low growth probability forecasts using a rolling-window of width 84 monthly observations. The binary stagnation series is based on the ARNG indicator.

Figure 12: U.S. and JAP out-of-sample *low growth* probabilities from both the dynamic binary and dynamic ordered probit models for different forecast horizons.

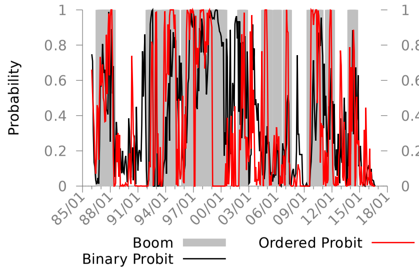
Panel A: U.S.



(a) 1-month forecast

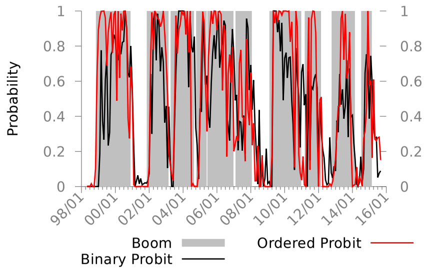


(b) 3-month forecast

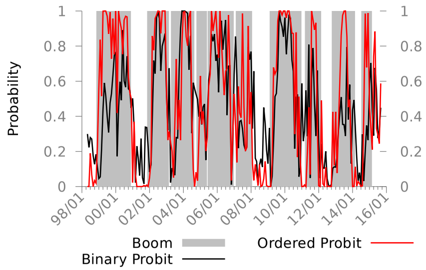


(c) 6-month forecast

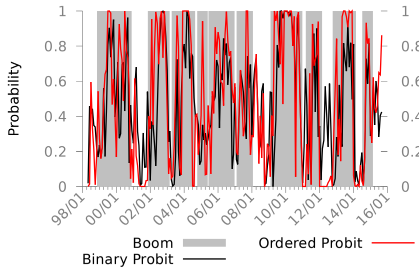
Panel B: JAP



(d) 1-month forecast



(e) 3-month forecast



(f) 6-month forecast

NOTES: These are the out-of-sample acceleration probability forecasts using a rolling-window of width 84 monthly observations. The binary acceleration series is based on the ARNG indicator.

Figure 13: U.S. and JAP out-of-sample *acceleration* probabilities from both the dynamic binary and dynamic ordered probit models for different forecast horizons.

Table 6: Test on *Area Under the ROC*-equality of out-of-sample *acceleration* probability forecasts between the ordered-probit and the binary-probit model by horizon.

	Test(USA)	p-value(USA)	Test(JAP)	p-value(JAP)
(A) Full sample				
h=1	0.117	0.732	15.244	0.000
h=2	0.131	0.718	27.035	0.000
h=3	1.052	0.305	28.072	0.000
h=4	0.797	0.372	27.994	0.000
h=5	0.457	0.499	26.376	0.000
h=6	0.364	0.546	15.953	0.000
(B) Since 2007m1 only				
h=1	14.301	0.000	9.509	0.002
h=2	19.866	0.000	17.083	0.000
h=3	19.917	0.000	14.666	0.000
h=4	15.935	0.000	12.167	0.000
h=5	12.673	0.000	12.107	0.001
h=6	17.364	0.000	7.209	0.007

Table 7: Test on *Area Under the ROC*-equality of out-of-sample *low growth* probability forecasts between the ordered-probit and the binary-probit model by horizon.

	Test(USA)	p-value(USA)	Test(JAP)	p-value(JAP)
(A) Full sample				
h=1	2.264	0.132	3.497	0.061
h=2	0.724	0.395	12.397	0.000
h=3	0.965	0.326	6.851	0.009
h=4	0.000	0.987	2.274	0.132
h=5	0.201	0.654	0.612	0.434
h=6	0.529	0.467	1.025	0.311
(B) Since 2007m1 only				
h=1	3.026	0.082	1.627	0.202
h=2	0.326	0.568	6.270	0.012
h=3	0.689	0.406	8.059	0.005
h=4	0.739	0.390	5.627	0.018
h=5	0.637	0.425	0.048	0.827
h=6	0.595	0.441	0.011	0.918

Impressum

Publisher: Hans-Böckler-Stiftung, Hans-Böckler-Str. 39, 40476 Düsseldorf, Germany
Phone: +49-211-7778-331, IMK@boeckler.de, http://www.imk-boeckler.de

IMK Working Paper is an online publication series available at:
http://www.boeckler.de/imk_5016.htm

ISSN: 1861-2199