International Journal of Forecasting 30 (2014) 898–917

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Predicting recessions with a composite real-time dynamic probit model

Christian R. Proaño[a,*], Thomas Theobald[b,1]

[a] The New School for Social Research, New York, NY, United States

[b] Macroeconomic Policy Institute (IMK), Düsseldorf, Germany

ARTICLE INFO

*Keywords:*
Dynamic binary choice models
Out-of-sample forecasting
Yield curve
Real-time econometrics

ABSTRACT

In this paper we propose a composite indicator for real-time recession forecasting based on alternative dynamic probit models. For this purpose, we use a large set of monthly macroeconomic and financial leading indicators from the German and US economies. Alternative dynamic probit regressions are specified through automated general-to-specific and specific-to-general lag selection procedures on the basis of slightly different initial sets. The resulting recession probability forecasts are then combined in order to decrease the volatility of the forecast errors and increase their forecasting accuracy. This procedure features not only good in-sample forecast statistics, but also good out-of-sample performances, as is illustrated using a real-time evaluation exercise.


## 1. Introduction

As is widely acknowledged, the timely and accurate prediction of turning points in the business cycle is one of the most policy-relevant aspects of macroeconomic forecasting. However, this task is also one of the most challenging. Not only are there many potential nonlinearities at the onset of a turning point in economic activity, there is also a significant level of uncertainty around macroeconomic data at the current edge,[2] in addition to the model uncertainty inherent in all applied work.

With respect to mitigating the model uncertainty problem, Bates and Granger (1969) were among the first to propose a combinatorial approach. They showed that the inclusion of inferior ex-ante forecasts could increase the predictive power of the best ex-ante forecasts if the inferior forecasts contained some novel information. More recently, Timmermann (2006) also emphasized the usefulness of forecast combination due to (1) diversification, (2) structural breaks, (3) misspecification of individual forecasts, and (4) systematic differences in the individual loss functions.

In contrast, methods for reducing the uncertainty inherent in end-point data are less developed. Pesaran and Timmermann (2005) have stressed the urgent need to develop robust interactive systems of model specification and evaluation which are designed explicitly to work in real time, as "by setting out in advance a set of rules for observation windows and variable selection, estimation, and modification of the econometric model, automation provides a way to reduce the effects of data snooping and facilitates learning from the performance of a given model when applied to a historical data set" (Pesaran & Timmermann, 2005, p. 212).

Binary response models have been used extensively in the literature for the prediction of business cycle turning

* Corresponding author.
*E-mail addresses:* christian.proano@gmail.com (C.R. Proaño), thomas-theobald@boeckler.de (T. Theobald).

1 Also Free University (FU) Berlin, Germany.

2 The current edge is defined as the last observation(s) of a certain vintage of macroeconomic data. These observations are usually subject to future data revisions. They are also called end-point data.

http://dx.doi.org/10.1016/j.ijforecast.2014.02.007
0169-2070/

points (Bernand & Gerlach, 1998; Chen, Iqbal, & Lai, 2011; Estrella & Mishkin, 1998; Estrella, Rodriguez, & Schich, 2003; Haltmeier, 2008; Hao & Ng, 2011; Moneta, 2005; Ng, 2012; Rudebusch & Williams, 2009; Wright, 2006). Along these lines, we discuss the rationale and structure of a composite indicator for real-time recession forecasting based on alternative dynamic probit models specified through automated *general-to-specific* and *specific-to-general* variable and lag selection procedures. This approach is specifically designed to work under real-time conditions, as was discussed by Proaño (2010).

Thus, the main contribution of this paper is the development of a composite dynamic probit indicator along the lines of recent studies using binary response models, such as those of Kauppi and Saikkonen (2008) and Nyberg (2010) for monthly recession forecasting under real-time conditions. As will be discussed in this paper, the estimation of several dynamic probit regressions and the combination of the resulting recession probability estimates takes into consideration the information from additional leading indicators, and achieves a higher recession forecast accuracy.

The remainder of this paper is organized as follows. In Section 2, we discuss in detail the structure of the composite real-time dynamic probit model and its underlying combination scheme. In Section 3, the real-time in- and out-of-sample performances of the composite model for the German and US economies are presented. A comprehensive comparison of our model with other existing approaches is conducted in Section 4. Finally, Section 5 draws some conclusions from this study and points out possible extensions for future research.

## 2. Methodology

As has been mentioned, following the work of Estrella and Hardouvelis (1991), binary response models have been used widely for the estimation and forecasting of recessionary periods over the last twenty years (Dueker, 1997; Kauppi & Saikkonen, 2008; Nyberg, 2010; Rudebusch & Williams, 2009). In this strand of the literature, the binary recession indicator series $b_t$, which represents the state of the economy within the business cycle, is set such that

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

Let $\Omega_{t-h}$ be the information set available at $t-h$, where $h$ represents the forecasting horizon. Assuming a one-period-ahead forecast horizon $h = 1$, $E_{t-1}$ and $\Pr_{t-1}(\cdot)$ denote the conditional expectation and the conditional probability given the information set $\Omega_{t-1}$. Under the assumption that $b_t$ has a Bernoulli distribution conditional on $\Omega_{t-h}$, i.e.

$$b_t |\ \Omega_{t-1} \sim \mathcal{B}(p_t),$$

the conditional probability $p_t$ of $b_t$ taking the value 1 in $t$ is given by

$$E_{t-1}(b_t) = \Pr_{t-1}(b_t = 1) = p_t = \Phi(E(\varphi_{t|t-1})),$$

where $\varphi_t$ represents a linear combination of the random variables contained in the information set $\Omega_{t-1}$. $\Phi(\cdot)$ represents the linking function between $\varphi_t$ and the conditional probability $\Pr_{t-1}(b_t = 1)$ according to the Bernoulli distribution, which, in probit models, is given by the standard normal distribution function.

The latent variable of the real-time dynamic probit indicator at hand is explained by various lags of the autoregressive reference series and a set of exogenous macroeconomic and financial leading indicators (which we discuss in detail below), summarized in the matrix $\mathbf{x}_t$, i.e.

$$\varphi_t = \sum_{j=h+D_y}^{p} \alpha_j y_{t-j} + \sum_{j=h+D_{\mathbf{x}}}^{q} \mathbf{x}'_{t-j}\beta_j + u_t,$$
$$u_t \sim N(0, 1)\ \forall t,\ R > D_y, \tag{1}$$

where $D_y$ and $D_x$ stand for the real-time data availability constraints.[3]

It should be clear that the inclusion of a large set of variables in $\mathbf{x}_t$ may lead to a serious multicollinearity problem if some series are highly correlated with others. This is likely to be the case if interest rates of government bonds at different maturities (or their spreads vis-à-vis the short-term interest rate) are included in $\mathbf{x}_t$ at the same time. In order to avoid this problem, we consider different specifications represented by $\mathbf{z}_t^i$ (the matrix which contains all of the explanatory variables of that particular specification), $i \in I$. Accordingly, the $i$th specification of the $h$-step-ahead recession forecast of the probit model regression is given by

$$\varphi^i_{t+h} = \mathbf{z}^{i'}_{\mathbf{t}}\beta + u^i_{t+h}, \quad u^i_{t+h} \sim N(0, 1),$$
$$i \in I,\ \text{with } b^i_{t+h} = \begin{cases} 1 : \varphi^i_{t+h} > 0 \\ 0 : \varphi^i_{t+h} \le 0, \end{cases} \tag{3}$$

where the size of $I$ is equal to the product of the combinatorial dimension and the elements in each of its components. For instance, with five different interest rate spreads and two different lag selection procedures, ten specifications can be taken into account.

Furthermore, in order to avoid the latent problem of choosing an arbitrary model specification based on an ad-hoc selection of lagged values – and of the explaining variables in general – each alternative dynamic probit specification is estimated using both a *general-to-specific* ($G$) and a *specific-to-general* ($S$) approach, following Proaño (2010). In the *general-to-specific* selection procedure

[3] Proaño (2010) explained the latent variable using various lags of the lagged binary variable $b_t$ in addition to the lagged reference series, i.e.,

$$\varphi_t = \sum_{j=h+R}^{o} \delta_j b_{t-j} + \sum_{j=h+D_y}^{p} \alpha_j y_{t-j} + \sum_{j=h+D_{\mathbf{x}}}^{q} \mathbf{x}'_{t-j}\beta_j + u_t,$$
$$u_t \sim N(0, 1)\ \forall\ t,\ R > D_y, \tag{2}$$

where $R$ stands for the recession recognition lag. Although we find that the inclusion of lags with this latter variable improves the out-of-sample real-time forecast accuracy slightly (see Appendix A), the inclusion of both autoregressive terms may produce multicollinearity problems. Following the advice of one referee, we only include the lagged values of the reference series $y_t$.