Bank of Finland

BoF Economics Review

7 • 2020

Probit based time series models in recession forecasting – A survey with an empirical illustration for Finland

Wilma Nissilä[1]

Abstract

This article surveys both earlier and recent research on recession forecasting with probit based time series models. Most studies use either a static probit model or its extensions in order to estimate the recession probabilities, while others use models based on a latent variable approach to account for nonlinearities. Many studies find that the term spread (i.e, the difference between long-term and short-term yields) is a useful predictor for recessions, but some recent studies also find that the ability of spread to predict recessions in the Euro Area has diminished over the years. Confidence indicators and financial variables such as stock returns seem to provide additional predictive power over the term spread. More sophisticated models outperform the basic static probit model in various studies. An empirical analysis made for Finland strengthens the findings of earlier studies. Consumer confidence is especially useful predictor of Finnish business cycle and the accuracy of the static single-predictor model can be improved by using multiple predictors and by allowing the dynamic extension.

Keywords: business cycles, recession forecasting, probit models

[1]*Department of Economics, University of Turku. Email: wilma.v.nissila@utu.fi. At the time of writing this article the writer was working at the Monetary Policy and Research Department of the Bank of Finland (May 2019 - September 2019).*

Disclaimer: The views and opinions expressed in this article are of the author and do not necessarily reflect the views of the Bank of Finland. I am grateful to Forecasting Unit of Bank of Finland, Juha Kilponen, Heikki Kauppi, Henri Nyberg and the participants of Nowcasting methods and Finnish Business Cycle Seminar for helpful comments and advise. I also thank Tomi Kortela for providing me the shadow rate estimates. Profound and valuable comments of the referee are appreciated. Part of this work was carried out within the financial support from Yrjö Jahnsson Foundation, Turun Kauppaopetussäätiö and TOP-Säätiö.

Probit Based Time Series Models in Recession Forecasting - A Survey with an Empirical Illustration for Finland[†]

Wilma Nissilä[‡]

August 7, 2020

Abstract

This article surveys both earlier and recent research on recession forecasting with probit based time series models. Most studies use either a static probit model or its extensions in order to estimate the recession probabilities, while others use models based on a latent variable approach to account for nonlinearities. Many studies find that the term spread (i.e, the difference between long-term and short-term yields) is a useful predictor for recessions, but some recent studies also find that the ability of spread to predict recessions in the Euro Area has diminished over the years. Confidence indicators and financial variables such as stock returns seem to provide additional predictive power over the term spread. More sophisticated models outperform the basic static probit model in various studies. An empirical analysis made for Finland strengthens the findings of earlier studies. Consumer confidence is especially useful predictor of Finnish business cycle and the accuracy of the static single-predictor model can be improved by using multiple predictors and by allowing the dynamic extension.

Keywords: Business Cycles, Recession Forecasting, Probit Models

[†]Disclaimer: The views and opinions expressed in this article are of the author and do not necessarily reflect the views of the Bank of Finland. I am grateful to Forecasting Unit of Bank of Finland, Juha Kilponen, Heikki Kauppi, Henri Nyberg and the participants of *Nowcasting methods and Finnish Business Cycle Seminar* for helpful comments and advise. I also thank Tomi Kortela for providing me the shadow rate estimates. Profound and valuable comments of the referee are appreciated. Part of this work was carried out within the financial support from Yrjö Jahnsson Foundation, Turun Kauppaopetussäätiö and TOP-Säätiö.

[‡]Department of Economics, University of Turku. Email: wilma.v.nissila@utu.fi. At the time of writing this article the writer was working at the Monetary Policy and Research Department of the Bank of Finland (May 2019 - September 2019).

1 Introduction

Fluctuations of economic activity are a central topic both in theoretical and empirical macroeconomic research. Predicting business cycles is one of the most challenging yet also one of the key activities performed by several economic institutions. Information about the future state of the economy is essential for policy makers, central banks, financial system surveillance authorities as well as private sector decision makers such as commercial banks and households. The importance of predicting the state of the economy is based on the fact that business cycles, especially recession periods, are costly. Since recession periods can be severe and long-lasting, many institutions try to mitigate the fluctuations of economic activity to the best of their knowledge. For example, government authorities can adjust spendings or central banks can review their monetary policy. However, if predictions are inaccurate, the timing of these policy actions may fail.

Therefore there has been a vital interest in tools for forecasting the economic activity. The literature on economic forecasting can be broadly divided into two branchs: one focusing on predicting the growth rates of GDP or other quantitative measures and the other trying to forecast the states of business cycles, i.e the recession and expansion periods of the economy. In macroeconomic research, so called probit models have been a standard tool to predict the business cycle states ever since the study of Estrella & Hardouvelis (1991). Estrella & Hardouvelis were the first authors, who in contrast to the previous literature[1], considered predicting the U.S recession periods with a static probit model. Various subsequent studies (see, e.g Estrella & Mishkin, 1998; Bernard & Gerlach, 1998 and Estrella, Rodrigues & Schich, 2003 among others) have since then employed the static probit model, but the past twenty years have also seen a significant amount of research on developing the baseline model (see, e.g Dueker, 1997; Chauvet & Potter, 2005 and Kauppi & Saikkonen, 2008 among others).

The literature on recession forecasting has identified several financial and sentiment-based variables as leading indicators of recession periods. These variables include term spread, stock returns and short-term interest rates, confidence indicators and credit variables. Much of the previous research lends support to the term spread being the main leading indicator (see, e.g Estrella & Mishkin, 1998 and Nyberg, 2010) although some recent studies argue that the predictive power of the term spread has diminished over the years in the Euro Area (see, e.g Fendel et al. 2018 and Pönkä & Stenborg, 2019) or that the term spread is not a reliable predictor for all countries (see, e.g Bernard & Gerlach, 1998 and Ahrens, 2002). In addition to identifying the potential predictive variables, the research on recession forecasting has paid attention to the accuracy of probit based models. Extended methods outperform the baseline model in various studies (see, e.g Chauvet & Potter, 2005; Kauppi & Saikkonen, 2008 and Ng, 2012).

The main purpose of this article is to survey the literature of probit based recession models. Earlier surveys (see, e.g Stock & Watson, 2003 and Wheelock & Wohar, 2009) have mainly concentrated on summarizing the research investigating either the predictive power of term spread or the ability of other financial variables to predict the output growth. Literature related to the forecasting recessions is interest and important for several reasons. First and most obviously, those who produce economic forecasts need to know which variables seem to provide reliable information

[1]Early work of recession forecasting includes the studies of Stock & Watson (1989) and Diebold & Rudebusch (1989). These studies did not apply probit models but alternative methods. Stock & Watson constructed the recession index by using dynamic factor models, whereas Diebold & Rudebusch estimated recession probabilities from the leading economic index with a Bayesian sequential probability recursion.

about the future state of the economy. Second, the literature survey is useful for those who are interested in understanding the mechanisms of the business cycles. Third, probit models can be estimated also by using the monthly data. Policy makers and central banks need as real-time data as possible and hence the views on the economic situation based solely on quarterly GDP are insufficient. GDP is considered to be the most central variable in determining the state of the economy, but the first GDP estimates for the previous quarter are published within a delay and usually revised significantly. This highlights the fact that higher frequency data is needed in macroeconomic forecasting.

Second purpose of this paper is to apply the static probit model and its extensions to predict the business cycle states in Finland. The majority of previous academic research on recession forecasting has been made for the United States or other large countries such as Germany. There exists rather little probit model related research on small and open economies such as Finland. At the time of writing this article, to the best of my knowledge, the only probit analysis of Finnish economy is by Pönkä & Stenborg (2019). Pönkä & Stenborg applied probit models for quarterly data examining the predictive power of five different variables. In this article, monthly data is used instead as well as a larger set of explanatory variables is considered.

This paper begins in section 2 in which a definition of a recession is presented and business cycle dating procedures are discussed. Next a brief introduction of the probit models is given emphasizing the developments made during the past twenty years. After that the criteria for evaluating the accuracy of these models and predictive variables are presented. Section 4 contains the review, which covers altogether 25 articles and working papers. Finally, the empirical analysis of recession periods in Finland is presented.

The results of this survey and illustration can be summarized as six main conclusions. First, various financial and sentiment based variables are useful predictors of recession periods. Term spread seems to be the most dominant predictor for forecasting U.S recessions, but the same does not apply for all considered countries or the Euro Area. Second, extensions of the static probit model outperform the baseline model. However, it remains unclear which extension yields the most accurate predictions, since the research results are mixed depending the data sample and employed variables. Third, there has been four separate recessions in Finland in 1988–2019 according to business cycle dating algorithms. Fourth, the power of the term spread to predict these recessions has decreased in Finland in 2010's. Forecasts based solely on the term spread date correctly the great recession of 1990's and the financial crisis in 2008, but not the start of the two euro crises. Modifying the term spread by using shadow rates improves the forecasts only slightly. Fifth, several confidence indicators, such as the consumers expectations of the general economic situation in Finland, are extremely useful in predicting the Finnish business cycle. Sixth, the most accurate results for Finland are obtained by using multi-predictor models and dynamic extension. Our results concerning the Finnish business cycle are in line with previous studies, especially with Pönkä & Stenborg (2019).

2 Defining and Dating Recessions

Before presenting the models that are used for predicting the business cycle states, it is necessary to define what is meant by a "recession". Since GDP has been considered to be the most central indicator of the economic activity, it is natural that the definition of recession is usually based on the fluctuations of GDP. For example, the decline in GDP for at least two consecutive quarters (Two Quarters Rule) is regarded as a recession by many economic instutions. However, a broader definition of recession has been favored in the US, where the National Bureau of Economic Research (NBER, 2008) provides the most widely accepted definition:

"A recession is a significant decline in economic activity spread across the economy, lasting more than few months, normally visible in real GDP, real income, employment, industrial production, and wholesale-retail sales. A recession begins just after the economy reaches a peak of activity and ends as the economy reaches its trough. Between trough and peak, the economy is in an expansion."

Hence, in determining the dates of business cycle turning points, i.e peaks and troughs of economic activity, the Business Cycle Dating Committee of NBER examines several variables instead of any single indicator such as GDP. This is due to two reasons. First, NBER applies the business cycle definition originally suggested by Burns & Mitchell (1946), who stated that the cycles are co-movement of multiple individual macroeconomic variables, which determine the turning points. The duration of a cycle from peak to peak or trough to trough can be anything from one to more than ten years. Second, NBER places considerable emphasis on monthly indicators since it has traditionally maintained a monthly chronology of turning points[2].

NBER has been dating the turning points of the US economy for almost 90 years and since 1978, when the Business Cycle Dating Committee was established, there has not been any changes to previously-announced business cycle turning points[3]. The turning point dating methodology applied by NBER is informal in a sense that the Business Cycle Dating Committee has no fixed rule in determining the turning points. In other words, the Committee applies its own judgement on the above definitions of recessions and expansions when deciding the "official" business cycle states of the US economy. Today, the NBER business cycle chronology is considered as a benchmark of US recessions periods.

However, if one accepts the NBER dates of peaks and troughs, some issues remain to consider. Peaks and troughs are announced within a substantial delay, since the Committee waits long enough so that the existence of the announced turning points is not in doubt. The announcement process can last several months or even more than a year due to the revisions of the initial values of macroeconomic data. For example, the second recent turning point in the US occurred in June 2009 and the Committee announced it as late as in September 2010, over a year after. The most recent peak, that occurred in February 2020, was instead published after four months in June 2020. There remains also some ambiguity about the duration of recessions, since for example a peak month could be classified either as the last month of an expansion or the first month of a recession. Due to this unclarity, the Federal Reserve Bank of Saint Louis calculates the durations of recessions using three different methods: midpoint method (peak and trough months are included in the recession),

[2]Business cycle dates of NBER are available at www.nber.org/cycles.html

[3]Prior to 1978 there were some revisions.

peak method (peak month is included in the recession but trough month is not) and trough method (peak month is not included in the recession but trough month is)[4]. All of these methods have been used in the recession forecasting literature.

The determination of monthly business cycle states has gained increasing attention among other economic institutions, because NBER determines the turning points only for the US economy. Economic Cycle Research Institute (ECRI) applies the same judgement based turning point dating procedure than NBER and has identified peak and trough months for 21 advanced and emerging countries[5]. Austria, France, Germany, Italy, Russia, Spain, Switzerland, Sweden and UK represent the European countries that have been examined by ECRI. ECRI also maintains the US business cycle chronology, but it is an identical to that of NBER. Besides ECRI, Euro Area Business Cycle Dating Committee of the Centre for Economic Policy[6] (CEPR) has accepted the NBER definition of a recession and identified the peak and troughs of the Euro Area by the same judgement based method. The CEPR turning point chronology is maintained quarterly[7], since the Committee argues that European statistics are of uneven quality, long time series are not available and data definions differ across the countries, which makes it difficult to maintain monthly chronology. Unfornately, CEPR has not dated the turning points for individual countries and therefore the Committee emphasizes that the business cycle dates may differ for the Euro Area and its member states.

The NBER, ECRI and CEPR turning points are widely accepted and used in business cycle research. However, there have been several attempts by researchers to formalize the judgement based turning point procedure. According to Hamilton (2011) there are at least three reasons why the formalization attempts are beneficial: aforementioned publication lag of NBER, political reasons[8] and that the mechanization of the informal procedure can help to understand better the business cycles. The formalization attempts can be broadly classified either parametric, which usually refer to the seminal work of Hamilton (1989) or non-parametric procedures, which refer to the seminal work of Bry & Boschan (1971).

The presentation of the full methodology related to turning point formalization exceeds the aim of this paper and therefore just a few key points are noted (see, e.g the survey of Hamilton (2011) and Bry & Boschan (1971, chapter 2) for more details). Hamilton (1989) extended Markov Switching models (MS models) to autoregressive processes and noticed that they can reproduce quite accurately the US business cycle states defined by NBER. Markov Switching models are a class of regime switching models, in which the behaviour of time series is expected to be different in two or more regimes. In recession forecasting literature, the number of regimes has been usually set to two to match the expansion and recessions periods[9]. The switching mechanism from one regime to another is controlled by unobservable state variable that follows a Markov chain. However, MS models do not define the exact turning points, but instead the (filtered and/or smoothed) regime probabilities, i.e the probabilities that the time series is in spesific regime at a given time. In order to classify the dates as recessions and expansions, some threshold value must be selected. Hamilton

[4]Recession series are available at fred.stlouisfed.org

[5]Turning point chronologies of ECRI are available at www.businesscycle.com

[6]CEPR and Euro Area Business Cycle Network have partnered from May 2019 to jointly support the work of Euro Area Business Cycle Dating Committee.

[7]The chronology of CEPR is available at cepr.org

[8]According to Hamilton (2011) "*. . . there is undeniably pressure to delay the announcement that a recession has has begun or accelerate the announcement that a recovery has begun if one's goal is to help the incumbent*".

[9]Sometimes three or four regimes are allowed. In three regime case, the economy is usually expected to have recession periods as well as high and low growth rate eras.

(1989) applied the 50 % threshold value, which is a natural choice but of course leaves a question of how probabilities near 50 % should be treated. In his original MS model, Hamilton allowed two regimes and the regime switching in the mean, but the model has been since then extended in many ways. For example, Hansen (1992) allowed the regime switching in parameters other than the mean, whereas Clements & Krolzig (1998) allowed more than two regimes and Chauvet (1998) combined the dynamic factor models with MS approach.

Non-parametric methods provide instead the exact turning point dates. Bry & Boschan (1971) developed an algorithm that comes closest to translating the NBER peaks and troughs into practise and can be seen as an extension of Two Quarters Rule. BB-algorithm is based on smoothing a univariate time series with moving averages and then extracting local minimum and maximum values from the series. Once the potential turning point candidates have been selected, they are filtered according to several restrictions: a cycle from peak to peak has a duration of 15 months, recessions and expansions last at least 5 months, identified peaks and troughs must alternate and the turning points can not be detected in real time, but 6 months after the turn has occurred (see Bry & Boschan, 1971, chapter 2 for more details). BB-algorithm is applied for monthly data that is seasonally adjusted but not trend eliminated. Hence the turning points do not depend on the selection of detrending method, which can be considered as a major advantage[10]. Harding & Pagan (2002) have provided the quarterly version of BB-algorithm, known as BBQ-algorithm, which omitted the smoothing procedure but maintained the restrictions and other principles of the original algorithm.

BB- and BBQ-algorithms have turned out to be quite succesful, but they do also suffer from shortcomings. First, the turning points are detected several months after the peak or trough has occurred. Second, BB- and BBQ-algorithm treat recessions and expansions symmetrically without taking into account the differences in average durations and drifts of different states. As a consequence the procedures may identify expansions phases that are both short and flat. To avoid this, Mönch & Uligh (2005) proposed an augmention to the original BB-algorithm by adding an amplitude/phase-length criterion that excludes too short and flat expansion periods. They define a ”short” expansion as a one whose length is outside the one-standard deviation interval around the average expansion length whereas the ”flat” expansion is a one in which the annualized growth rate is outside the one-standard deviation interval around the average positive annual growth rate. Third, Morley & Piger (2005) noticed that BBQ-algorithm makes some systematic errors when selecting the turning points of the US and suggested a slight modification at the first step of the algorithm[11]. The modified BBQ-algorithm was found to be more accurate than the original. Fourth, even the extended and modified BB- and BBQ-algorithms are applicable only for the univariate time series which leaves two options. Either BB-algorithm is applied to individual coincident series[12] and individual turning points are aggregated or a coincident index is constructed and BB-algorithm is applied to the index instead of individual series. For turning point aggregation, Harding & Pagan

[10]Dating methods that are based on detrending the series are sometimes called *growth cycle dating methods*, whereas identifying the turning points from the level of the series is called *classical business cycle dating*. This survey focuses on the classical view.

[11]In the heart of the first step is a definition of a local maxima (minima) as occuring at time t when $\{y_t$ - $y_{t\pm k} >$ $(<)\ 0\}$, $k = 1\ , \ldots,\ K$ where K is set to five or two depending the frequency of series y_t. Morley & Piger replaced the zeros by threshold values.

[12]Coincident variables such as GDP, employment or industrial production change approximately at the same time as the whole economy thus providing information about the current state of the economy.

(2006) have developed an algorithm that clusters the turning points of individual series, while Stock & Watson (2014) provide an alternative method.

The accuracy of parametric and non-parametric methods has been compared rather little in the recent research. Harding & Pagan (2003a) compared the BBQ-algorithm and Hamilton's (1989) MS model with univariate US data and found that results achieved with MS model may lack stability when sample size or model changes. Therefore they ended up recommending to use the BB-algorithm or its modifications[13]. Ahking (2014) compares original BB-algorithm and Hamilton's (1989) MS model with US data and finds very similar results than Harding & Pagan (2003a). Ahking concludes that MS model is not robust to the sample period and to a slight change in the variable used in determining the turning points. Thus Ahking ends up also recommending to use parametric methods in business cycle research. Chauvet & Piger (2008) instead compared the aggregating method of Harding & Pagan (2006) and Markov Switching dynamic factor model of Chauvet (1998) with real-time multivariate US data and found that MS procedure was slightly more accurate in dating US business cycles. However, they also stated that "... *both approaches are capable of identifying turning points in real time with reasonable accuracy.*" General consensus of superiority of one dating method over the other has not been reached. Therefore some researchers have used both methods in turning point dating (see, e.g Bengoechea & Perez Quiros, 2004; Bruno & Otranto, 2008; Schirwitz, 2009; among others).

[13] See the response of Hamilton (2003) and the counterpart of Harding & Pagan (2003b).

3 Probit Based Time Series Models in Recession Forecasting

3.1 Static Probit Model and its Extensions

The dependent variable of a probit model is defined as a binary indicator

$$y_t = \begin{cases} 1, \text{if the economy is in a recession at time } t \text{ and} \\ 0, \text{if the economy is in an expansion at time } t. \end{cases}$$

We are interested in modelling of y_t by using different explanatory variables. Let Ω_{t-1} be the the information set available at time $t-1$ and assume that, conditional on Ω_{t-1}, y_t follows a Bernoulli distribution

$$y_t|\Omega_{t-1} \sim B(p_t), \tag{1}$$

where the conditional probability, p_t, is according to the properties of Bernoulli distribution

$$p_t = E_{t-1}(y_t) = P_{t-1}(y_t = 1) = \Phi(\pi_t). \tag{2}$$

In equation (2) E and P signify the conditional expectation and conditional probability given the information set Ω_{t-1}, whereas Φ is the cumulative distribution function of a continuous random variable and π_t is a linear function of explanatory variables included in the information set Ω_{t-1}. In practise, the function Φ is usually assumed to be the cumulative distribution function of a standard normal distribution or a logistic distribution. The former assumption leads to a probit model, whereas the latter leads to a logit model[14]. In this article, we focus on probit models.

In a static probit model the function π_t is specified as

$$\pi_t = \omega + x'_{t-k}\beta, \tag{3}$$

where ω is a constant and $x_{t-k}, k \geq 1$ is a vector containing the explanatory variables. Lagged values of explanatory variables may also be included in x_{t-k}. Specification (3) is "static" in a sense that explanatory variables have an immediate effect on the p_t, since p_t does not change unless the values of x_{t-k} change.

A significant limitation of the static probit model is that it does not take into account the potential autocorrelation in dependent variable y_t. Thus, an obvious extension to the static model (3) is to include a value of $y_{t-l}, l \geq 1$ as an explanatory variable. For example, Dueker (1997) and Nyberg (2010), among others, have considered the dynamic probit model

$$\pi_t = \omega + x'_{t-k}\beta + \delta_1 y_{t-l}. \tag{4}$$

In model (4), only one lag of y_t is included, but it is possible to generalize the model by including any number of lags. In practise, when recessions are predicted in real time, one must however note that the values of the binary indicator y_t are known with a delay. Dueker (1997) argues that three months is probably a minimum recognition lag time for recessions[15]. Kauppi & Saikkonen (2008) extended the model (4) into the dynamic autoregressive probit model

$$\pi_t = \omega + x'_{t-k}\beta + \delta_1 y_{t-l} + \alpha_1 \pi_{t-1}, \tag{5}$$

[14] In empirical applications, probit and logit models yield very similar results (see, e.g Maddala, 1983, 23).

[15] According to Dueker (1997), it would not be reasonable to include last month's value of the recession binary variable as an explanatory variable, since it takes more time to recognize that the economy is in a recession. Dueker also argues that even if NBER may not have officially announced a turning point at three months, there has been other information for decision makers to infer whether the economy is in a recession or not.

where the condition $|\alpha_1| < 1$ is assumed. Again, for simplicity, only one lag of π_t is included in (5), but the model can be generalized by adding several lags of π_t. By using recursive substitution and the condition $|\alpha_1| < 1$, model (5) can be expressed as

$$\pi_t = \sum_{i=1}^{\infty} \alpha_1^{i-1}\omega + \sum_{i=1}^{\infty} \alpha_1^{i-1} x'_{t-k-i+1}\beta + \delta_1 \sum_{i=1}^{\infty} \alpha_1^{i-1} y_{t-l-i+1}.$$

Thus, π_t can be expressed in terms of the infinite history of y_t and the explanatory variables x_t. Therefore, in cases where many lags of dependent variable and explanatory variables are needed in specification (4), dynamic autoregressive model may provide more parsimonious and hence, preferable alternative. The autoregressive probit model

$$\pi_t = \omega + x'_{t-k}\beta + \alpha_1 \pi_{t-1} \tag{6}$$

is obtained from (5) by restricting the coefficient δ_1 to zero.

Traditional probit model and its extensions use typically only a few explanatory variables. To account for the large number of possible predictors, the models of the form (3)-(6) have also been augmented by using factors as predictive variables (see, e.g Chen et al. 2011, Bellégo & Ferrara, 2012 and Christiansen et al. 2014). These models are referred to as factor-augmented probit models in this article.

3.2 Latent Variable Approach

An alternative way to specify a probit model is based on a latent variable approach. In this approach, it is assumed that the dependent variable y_t is defined as

$$y_t = \begin{cases} 1, \text{if } y_t^* > 0 \text{ and} \\ 0, \text{if } y_t^* \leq 0, \end{cases}$$

where y_t^* is a continuous variable that is not observable, i.e a latent variable. The static probit model (3) can be based on the equation

$$y_t^* = \omega + x'_{t-k}\beta + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d}(0, 1). \tag{7}$$

The normality assumption of ε_t in (7) leads to a static probit model

$$\begin{aligned} p_t = P_{t-1}(y_t = 1) &= P_{t-1}(y_t^* > 0) \\ &= P_{t-1}(\omega + x'_{t-k}\beta + \varepsilon_t > 0) \\ &= P_{t-1}(\varepsilon_t > -\omega - x'_{t-k}\beta) \\ &= 1 - P_{t-1}(\varepsilon_t < -\omega - x'_{t-k}\beta) \\ &= 1 - \Phi(-\omega - x'_{t-k}\beta) \\ &= \Phi(\omega + x'_{t-k}\beta), \end{aligned}$$

where P signifies again the conditional probability given the information set and Φ is the cumulative distribution function of a standard normal distribution. Note that the last equality follows by the symmetry of normal distribution.

The static probit model in latent variable approach has been extended in many ways. For example, Poirier & Ruud (1988) have considered a latent variable model where the error term ε_t

is expected to follow a first order autoregressive process. Chauvet & Potter (2005) have suggested various extensions to (7) to account for the possibility of multiple breaks and autocorrelation. In their dynamic version the dependent variable y_t^* follows a first order autoregressive process

$$y_t^* = \omega + x_{t-k}'\beta + \theta_1 y_{t-1}^* + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d}(0,1), \tag{8}$$

where the autoregressive parameter $|\theta_1| < 1$ and the error term is allowed to be time-varying by the variance term σ_t. Chauvet & Potter concluded that this more sophisticated model outperformed the standard static version.

3.3 Estimation and Forecasting

The parameters of probit models of the form (3)-(6) can be estimated by using the traditional maximum likelihood (ML) estimation method (see appendix A). In the case of the models (3)-(4), de Jong & Woutersen (2011) showed that, under appropriate regularity conditions, the conventional large sample theory of ML estimation applies. The assumed regularity conditions include, for example, the stationarity of explanatory variables and correctness of a probit specification. Under these conditions, the ML estimator is consistent and asymptotically normal. Extending this theory to the models with autoregressive components, such as (5)-(6), remains still to be done. However, the results of de Jong & Woutersen seem to indicate, that under reasonable conditions the ML estimator of a model containing an autoregressive term is also consistent and asymptotically normal. Assuming this, Kauppi & Saikkonen (2008) show how robust standard errors allowing for autocorrelation can be obtained.

Latent variable probit models can also be estimated by using the traditional ML estimation method. Estimating the parameters of the model (8) by ML method requires, however, multiple integration over the unobserved lagged variable. Therefore Chauvet & Potter (2005) used Bayesian method[16] based on the Gibbs sampler to estimate (8), but the needed computations were quite extensive.

Thus, a major advantage of observation-driven probit models is that their estimation can be carried out by standard numerical methods. Another important advantage is that the one- and multi-period forecasts can be computed from explicit formulas. This is in contrast to many other nonlinear time series models such as probit models based on a latent variable approach. Kauppi & Saikkonen (2008) proposed two methods, "direct" and "iterative", for computing the multiperiod forecasts. In direct setting, the forecasts can be obtained by directly of the right sides of the equations (3)-(6), whereas in iterative approach the computation is a bit more difficult. Iterative method applies the same one period model iteratively accounting for all possible paths and their probabilities (see appendix B and Kauppi & Saikkonen, 2008, for more details). The ranking between direct and iterative forecasting procedures depends whether the model used in iterative forecasting is close to or far from the true data generating process. If it is close, then iterated forecasts tend to be more efficient, but if it is far, then direct forecasts are superior. (Kauppi & Saikkonen, 2008.) Kauppi & Saikkonen compared both methods for the US data and concluded, that in this case, the iterative forecasts tended to be superior.

[16] Bayesian methods have become popular in econometrics, especially when models are used for real-time forecasting purposes. See, for example, Geweke & Whiteman (2006) for an overview.

3.4 Goodness-of-fit Measurement

Various goodness-of-fit measures have been proposed to evaluate the fitted values and forecasts obtained with probit models. One of the most commonly used criteria is a counterpart of the coefficient of determination (R^2), designed for binary response models. This counterpart is typically called a pseudo-R^2.

Estrella (1998) proposed the following pseudo-R^2 (ps.R^2)

$$ps.R^2 = 1 - \left(\frac{\log L_u}{\log L_c}\right)^{-(2/T)\log L_c},$$

where $\log L_u$ denotes the unconstrained maximum value of the log-likelihood function, $\log L_c$ denotes the corresponding maximum value when all coefficients but a constant are restricted to zero and T is the sample size. Pseudo-R^2 takes values in the unit interval. The endpoints of the interval correspond straightforwardly to "no fit" and to "perfect fit". Pseudo-R^2 does not, however, take into account the amount of explanatory variables used in a model. Therefore adjusted version of pseudo-R^2 is sometimes used as a goodness-of-fit measure, since it allows to compare the models containing a different amount of explanatory variables. Adjusted pseudo-R^2 is defined as

$$adj.ps.R^2 = 1 - \left(\frac{(1-R^2)(T-1)}{T-k-1)}\right),$$

where k is the number of the explanatory variables.

Third commonly used evaluation measure is the quadratic probability score QPS (see, Diebold & Rudebusch, 1989) that is defined as

$$\text{QPS} = \frac{1}{T}\sum_{t=1}^{T} 2(y_t - p_t)^2,$$

where y_t is a value of a binary indicator and p_t denotes the fitted probability. QPS takes values between 0 and 2, endpoints corresponding to "perfect fit" and to "no fit".

Signal predictions and succes ratios have also been used in probit model evaluation. A signal prediction of a binary indicator y_t can be written as

$$\hat{y}_t = I(p_t > \zeta)$$

where I is the indicator function and ζ is a threshold. If a fitted probability p_t is greater than the threshold ζ, the signal prediction is $\hat{y}_t = 1$ and if $p_t \leq \zeta$, then $\hat{y}_t = 0$. Success ratio SR is the percentage of correct signal predictions. In calculating the success ratio, some pre-determined threshold value must be selected. A choice $\zeta = 0.5$ is a commonly used and natural in probability sense, but it may not be an optimal choice. Recessions are rare compared to expansion periods, so the threshold may be time-varying or considerably lower than the natural threshold 0.5.

Since there is no rule on setting some specific threshold, success ratios may not be the most informative measures on evaluating the predictive accuracy of probit models. The Receiver Operating Characteristic (ROC) curve is an alternative method to asses the goodness-of-fit of binary response models. The ROC curve has become a traditional tool in medical applications and biostatistics, but it has also recently gained popularity among economic applications (see, for example Berge & Jorda, 2011 and Christiansen et al. 2014). The ROC curve is a mapping of the true positive rate

$$TPR(\zeta) = P(p_t > \zeta | y_t = 1)$$

and the false positive rate

$$FPR(\zeta) = P(p_t > \zeta | y_t = 0)$$

for all possible thresholds ζ, described as an increasing function in the [0,1] $\times$ [0,1] space with TPR(ζ) on the Y-axis and FPR(ζ) on the X-axis. A ROC curve above the 45-degree line indicates predictive accuracy superior to coin toss. Instead of graphical representation, the area under the ROC curve (AUC) is a more convenient way to summarize the predictive information contained in the ROC curve. The AUC is defined as the integral of the ROC curve between 0 and 1 and thus the AUC takes values in the unit interval.

Information criteria can also be used to determine the fit of a probit model. The most widely used choices are the criteria of Akaike (1974) and Schwarz (1978).

Aforementioned goodness-of-fit criteria can be calculated for probit models estimated over the entire sample period (in-sample results) and for models that are estimated by using the expanding window approach (out-of-sample results). Typically, more emphasis is given to the out-of-sample results, since they tend to give a more realistic view of the predictive ability of the model. However, it is worth noting that when calculating out-of-sample results, there is no guarantee that the value of the pseudo-R^2 will lie in the unit interval, as discussed in Estrella & Mishkin (1998).

4 The Literature Review

In this section we present the survey reviewing the earlier and recent probit model related literature. For the sake of brevity, we focus on probit models that are used for predicting recession periods. Our survey covers altogether 25 articles and working papers. Covered studies are represented in table 1 together with the information of the data sample used, countries investigated and the principal findings based on-out-sample results or in-sample results, if out-of-sample results were not available. We note that majority (15 out of 25) of these studies considers the U.S recession periods alone which complicates drawing conclusions. We note also that more research is needed, especially for smaller and non-industrialized countries, since nearly all previous studies have mostly focused on large industrialized economies.

Estrella & Hardouvelis (1991) were the first ones to apply a static probit model to predict the U.S recession periods. They find that the term spread between the yields on 10-year and 3-month Treasury securities is useful for forecasting recessions 4 quarters ahead and that the result is robust to including other variables in model. Estrella & Hardouvelis do not consider any other forecast horizons than 4 quarters, since the main scope of their article was to predict U.S GDP growth rate. After their seminal work, many subsequent studies have employed the static model and it has become standard to use the difference between 10-year and 3-month yields as an estimate of term spread.

Estrella & Mishkin (1998) consider a variety of horizons from 1 to 8 quarters and include a large set of financial variables as possible predictors. They find that the term spread is a dominant predictor at horizons beyond 2 quarters and that stock returns bring additional predictive power up to 3 quarters. Estrella & Mishkin also highlight that the great in-sample result does not guarantee the good performance in out-of-sample estimation. For example, the commercial paper spread and a leading indicator of Commerce Department performed great in in-sample estimation, but the deterioration in performance was substantial when the predictive ability was tested in pseudo out-of-sample setting. The focus of the article of Estrella & Mishkin is in the U.S recessions, but Bernard & Gerlach (1998) and Ahrens (2002) employ a static probit model for several industrialized countries. They conclude that term spread is useful for the most of the countries, but it has very limited predictive power in Japan, Netherlands, UK and Italy, which emphasizes that the term spread is not a dominant or even a useful predictor for all economies. Euro Area recessions were predicted by Moneta (2003) and Duarter, Venetis & Paya (2005), who find the term spread a useful predictor by using the data sample from 1970 to 2000–2002.

Estrella, Rodrigues & Schich (2003) study the stability of term spread as a predictor of the U.S and German recessions. They conclude that the spread is both a stable and a useful recession predictor, but Chauvet & Potter (2005) find opposite results. Camacho & Perez-Quiros (2002) apply probit model with a leading indicator of Conference Board and find that other nonlinear models predict recessions better. Wright (2006) address the additive predictive power of federal funds rate.

Dueker (1997) was the first to employ a dynamic extension to a probit model. Moneta (2003) and Duarter, Venetis & Paya (2005) consider also dynamic extensions. They all find that the dynamic component brings additional predictive power to the static model. Kauppi & Saikkonen (2008) provided an autoregressive probit model and extended the dynamic model with an autoregressive component to allow even more richer dynamics. However, the dynamic model was found to be the most accurate. Nyberg (2010) applied models of Kauppi & Saikkonen and allowed an interaction term between a dynamic component and the other explanatory variables exluding the autoregressive component. Kauppi & Saikkonen considered only the term spread as a predictor, but Nyberg tested various other financial variables ending up with the same conclusion as Estrella & Mishkin (1998), i.e stock returns bring additional power. Fornari & Lemke (2010) and Ng (2012) highlight also the role of financial variables in recession forecasting.

In previously mentioned studies it has been common to apply single-predictor models or multi-predictor models with only a few explanatory variables. Chen et al. (2011) augmented probit models by allowing factors, estimated by principal component analysis (PCA), to be explanatory variables. These factor-augmented models have been found to outperform the static probit model in various studies including Chen et. al, Bellégo & Ferrara (2012) and Fossati (2015).

It is quite interesting that until the study of Christiansen et al. (2014), the role of sentiment was given less emphasis in recession forecasting and the related research was mostly focusing on employing financial variables and extending the static model. However, Christiansen et al. (2014) show that especially consumer confidence index and Purchasing Managers Index (PMI) are extremely useful predictors and their predictive power is robust to including the classical predictors. As a measure of consumers confidence they use the University of Michigan's Index of Consumer Sentiment, which is based on a monthly survey, where a minimum of 500 households are interviewed about their financial situation and the general business conditions in the U.S. Consumer confidence represents the expectations of the demand side, while the expectations of the supply are measured by PMI of Institute of Supply Management. PMI is constructed through a survey of more than 400 industrial companies in 20 manufacturing industries across the US. The respondents answer questions that compare the current level of activity with that of the previous month. Karnizova & Li (2014) supplement the results of Christiansen et al. by concluding that economic policy uncertainty indices are also useful sentiment-based predictors.

More potential recession predictors were found by Pönkä (2017), who concludes that credit variables, especially excess bond premium, bring additional predictive power to various forms of probit models. When predicting the quarterly business cycle states of Finland, Pönkä & Stenborg (2019) find real house prices useful predictors. Pönkä & Stenborg also noted that the predictive ability of term spread has decreased in Finland during the years and the spread could not forecast the start of the euro crisis. Fendel et al. (2018) draw a same conclusion but for the Euro Area. However, the shadow rate modified version of the spread is showed to perform better (for shadow rates, see e.g Wu & Xia, 2016 and Kortela, 2016 and the discussion in Fendel et al.).

The covered studies mainly focus on the empirical aspect of the predictive ability of different variables and discuss much less why term spread, stock returns or sentiment based variables might be useful and stable predictors of future recessions. Theoretical explanation exceeds also the aim of this article and is left for future research. However, some key points are noted.

The so-called expectations hypothesis of the term structure is usually the foundation of most explanations of the term spread's usefulness. The expectations hypothesis states that long-term interest rates equal the sum of current and expected future short-term interest rates plus a term premium. Term premium explains the positive slope of term spread and typically, if consumers expect short-term rates to fall, term spread approaches to zero or turns to negative. Negative term spread has for example preceded nearly all U.S recession periods. Rosenberg & Maurer (2008) conclude that ethe xpectations component of term spread is a more useful predictor than a term premium when U.S recessions are considered.

Stock prices can be interpreted as expected discounted values of future divided payments. They incorporate consumers and investors expectations regarding both the profitability of the firm and future interest or discounting rates. Sentiment based variables in turn reflect the expectations of the future economic situation thus being procyclical in relation to business cycles.

Thus all previously mentioned variables are somehow expectations and confidence related. Ben Bernanke, a former Chairman of the Board of Governors of the Federal Reserve System, has stated[17] that

> *"As in all past crises, at the root of the problem is a loss of confidence by investors and the public in the strength of key financial institutions and markets. The crisis will end when comprehensive responses by political and financial leaders restore that trust, bringing investors back into the market and allowing the normal business of extending credit to households and firms to resume."*

Traditionally, more emphasis has been given to financial variables and interest rates in recession predicting, but the studies of Christiansen et al. (2014) and Pönkä & Stenborg (2019) demonstrated that the role of sentiment cannot be neglected. It seems that confidence indicators contain some information that is not captured by the standard classical predictors such as the term spread and stock returns. However classical variables remain relevant predictors, at least for the U.S recession periods. The literature survey also highlights the role of extended probit models.

[17] The speech is available at https://www.federalreserve.gov/newsevents/speech/bernanke20081015a.htm

Table 1: Selective summary of probit model studies.

Study	Model	Data	Principal findings	Notes
Estrella & Hardouvelis (1991)	Static probit	U.S, 1955–1988 (Q)	Term spread is useful for predicting recessions 4 quarters ahead and results are robust to including other variables in the model	Results are based on in-sample estimation
Dueker (1997)	Static probit	U.S, 1959–1995 (M)	Term spread is dominant predictor at horizons beyond 3 months	Results are based on in-sample estimation
	Dynamic probit	U.S, 1959–1995 (M)	Power of dynamic component worsens at longer horizons, spread remains dominant predictor	Results are based on in-sample estimation
	Markov Switching probit	U.S, 1959–1995 (M)	Markov switching is relevant at longer horizons, spread remains dominant predictor	Results are based on in-sample estimation
Bernard & Gerlach (1998)	Static probit	U.S, UK, JPN, GER, FR, BEL, CAN, NL 1972–1993 (Q)	Term spread is useful predictor in all countries, but it has very limited predictive power in Japan and Netherlands. Best results are obtained for Germany, US and Canada. Foreign spreads bring only limited additional information.	Results are based on in-sample estimation. Leading indicators were also considered as predictors, but they bring additional information only for very short horizons
Estrella & Mishkin (1998)	Static probit	U.S, 1959-1995 (Q)	Term spread is dominant predictor at horizons beyond 2 quarters, stock returns have additional predictive power up to 3 quarters	Various other financial variables were also considerd as predictors of recessions
Ahrens (2002)	Static probit	U.S, UK, JPN, GER, FR, IT, CAN, NL 1970–1996 (M)	Filtered regime probabilities from Markov switching model, estimated by using term spread, are useful predictors for all countries except Italy. Predictive power of probabilities is very limited for UK and Japan.	Results are based on in-sample estimation. On averare, the Markov switching filter does not improve the predictive power of spread.
Camacho & Perez-Quiros (2002)	Static probit	U.S, 1960–1997 (Q)	Leading indicator (LEI) of Conference Board is a useful predictor, but more accurate results are obtained with other nonlinear models	No other variables considered
Estrella, Rodrigues & Schich (2003)	Static probit	U.S and GER 1967–1998 (M)	Predictive power of term spread is stable. Term spread is most useful at 12-month forecast horizon compared to 24- and 36-month horizons	Results are based on in-sample estimation, no other variables considered
Moneta (2003)	Static probit	Euro Area 1970–2002 (Q)	Term spread is useful for predicting recessions 6 quarters ahead	Best result is obtained at horizon of 4 quarters
	Dynamic probit	Euro Area 1970–2002 (Q)	Dynamic probit outperforms static model	Best result is obtained at horizon of 4 quarters
Chauvet & Potter (2005)	Latent variable approach	U.S, 1954–2000 (M)	Dynamic version with time varying error term outperforms the baseline model	No other variables considered. Predictive power of spread is not stable.
Duarte, Venetis & Paya (2005)	Static probit	Euro Area 1970–2000 (Q)	Term spread is useful for predicting recessions 3 quarters ahead U.S spread has additional predictive power	
	Dynamic probit	Euro Area 1970–2000 (Q)	Dynamic probit outperforms static model	

Notes: In the table Q and M denote the quarterly and monthly frequency of data. Findings are based on out-of-sample estimation results if not mentioned otherwise.

Selective summary of probit model studies (Table 1 continues).

Study	Model	Data	Principal findings	Notes
Wright (2006)	Static probit	U.S, 1964–2005 (Q)	Model with term spread and federal funds rate yields better results than model including spread only	
Kauppi & Saikkonen (2008)	Static probit	U.S, 1955–2003 (Q)	Term spread is useful predictor even at the longer forecast horizons	No other variables considered
	Dynamic probit	U.S, 1955–2003 (Q)	Dynamic probit out outperforms the static model	Publication lag of NBER assumed
	Autoregressive probit	U.S, 1955–2003 (Q)	Autoregressive probit outperforms static model but not dynamic	
	Dynamic autoregressive probit	U.S, 1955–2003 (Q)	Dynamic autoregressive probit outperforms static & autoregressive models, but not dynamic	Publication lag of NBER assumed
Rosenberg & Maurer (2008)	Static probit	U.S, 1961–2006 (M)	Expectations component of term spread is more useful than term premium component for predicting recessions 12 months ahead	Spread remains useful when federal funds rate is included in the model
Nyberg (2010)	Static probit	U.S and GER 1972–2007 (M)	Domestic term spread is dominant predictor for both countries Stock returns and foreign spread have additional predictive power	Interest rate differential is useful predictor for Germany
	Autoregressive probit	U.S and GER 1972–2007 (M)	Autoregressive probit outperforms static and dynamic model for both countries	Model with interaction term is also considered
	Dynamic probit	U.S and GER 1972–2007 (M)	Dynamic model outperforms the static model	Publication lag of NBER assumed
Fornari & Lemke (2010)	VAR-augmented probit	U.S, GER, JPN 1960–2009 (Q)	Term spread is a useful predictor, but various financial variables, such as stock returns and short term interest rates, bring additive predictional power	Good results are obtained for U.S and Germany, but the forecasts for Japan are considerably inferior
Chen, Iqbal & Lai (2011)	Factor-augmented probit	U.S, 1964–2007 (M)	Factor-augmented probit outperforms three static models: model including spread, model including spread and stock returns and model including spread and federal funds rate.	Eight factors are estimated by PCA by using a panel of 141 variables. Six of them are used in factor-augmented probit model.
Bellégo & Ferrara (2012)	Factor-augmented probit	Euro Area 1974–2008 (M)	Best results are obtained with a horizon of 12 months. Including lagged values of factors improves the in-sample fit.	Two factors are estimated from a panel of variables and then used in a model
Ng (2012)	Static probit	U.S, 1963–2010 (M)	Term spread is a useful predictor. TED spread, stock returns and composite index of eight leading indicators bring addictive power	Best results are obtained with a horizon of 4 months
	Dynamic probit	U.S, 1963–2010 (M)	Dynamic probit outperforms static model in all but one forecast horizons (5 months) by using 50% threshold	Publication lag of NBER assumed
	Autoregressive probit	U.S, 1963–2010 (M)	Autoregressive probit outperforms dynamic probit only in one forecast horizon (6 months) by using 50% threshold	
	Dynamic autoregressive probit	U.S, 1963–2010 (M)	Dynamic autoregressive probit outperforms autoregressive model, but not simple dynamic model by using 50% threshold	Publication lag of NBER assumed

Notes: In the table Q and M denote the quarterly and monthly frequency of data. Findings are based on out-of-sample estimation results if not mentioned otherwise.

Selective summary of probit model studies (Table 1 continues).

Study	Model	Data	Principal findings	Notes
Karnizova & Li (2014)	Static probit	U.S, 1985–2013 (Q)	Indices of economic policy uncertainty, especially the index constructed from newspaper reports related to economic policy uncertainty, are useful recession predictors	Result concerning news-index is robust to including the term spread
Christiansen, Eriksen & Moller (2014)	Static probit	U.S, 1978–2011 (M)	Two sentiment based variables, Consumer Confidence and Purchasing Managers Index, outperform spread with horizons less than 9 months	Sentiment variables bring also additive predictive power with longer horizons
	Factor-augmented probit	U.S, 1978–2011 (M)	Factor-augmented probit with two sentiment variables outperforms the static model with sentiment variables and spread only with forecast horizons under 3 months	15 factors are estimated by using a panel of 176 variables. Three of them are used in factor-augmented probit model
	Dynamic factor augmented probit	U.S, 1978–2011 (M)	Results concerning the sentiment variables are robust to including the dynamic component in models	Dynamic component is statistically either marginally significant or insignificant
Fossati (2015)	Factor-augmented probit	U.S, 1967–2005 (M)	Factor-augmented probit model outperforms static model that includes term spread, stock returns, federal funds rate and growth rate of non-farm employment	Three factors are estimated by using a panel of 30 variables
Fornaro (2016)	Latent variable approach with Bayesian shrinkage	U.S, 1959–2014 (M)	Bayesian shrinkage method outperforms static model with short-run forecast horizons and factor-augmented model with long-run forecast horizons	
Fendel, Mai & Mohr (2018)	Static probit	Euro Area 2001 – 2017 (M)	Predictive ability of term spread has diminished after euro crisis and modified version of spread produces more accurate results	Various variables like real M1, Purchasing Managers Index, Libor-OIS spread and crude oil are useful predictors
Pönkä (2017)	Static probit	U.S, 1973–2012 (M)	Credit spread index brings additional predictive power over term spread up to 3 months and excess bond premium up to 12 months	Results are robust to including stock returns, consumer confidence and federal funds rate
	Factor-augmented probit	U.S, 1973–2012 (M)	Static model with excess bond premium and classical predictors outperforms various factor-augmented model variants	17 factors are estimated by using a panel of 180 variables. Three of them are used in factor-augmented probit model.
	Autoregressive factor augmented probit	U.S, 1973–2012 (M)	Autoregressive component brings very little additive predictive power in some factor-augmented as well as static models	
Pönkä & Stenborg (2019)	Static probit	FI, 1988–2019 (Q)	Term spread, consumer confidence index and real house prices are dominant predictors from 1 to 4 quarters ahead. The predictive ability of term spread has diminished after euro crisis.	4-quarter changes in house prices and 4-quarter stock returns are more useful than quarterly changes or returns
	Autoregressive probit	FI, 1988–2019 (Q)	Autoregressive component is useful for some forecast horizons	

Notes: In the table Q and M denote the quarterly and monthly frequency of data. Findings are based on out-of-sample estimation results if not mentioned otherwise.

5 Empirical Analysis of Recession Periods in Finland

5.1 Business Cycle Chronology

One of the main issues when applying a probit model for Finland is the selection of the business cycle chronology, i.e the chronology of the binary indicator y_t discussed in section 3.1. Finland has witnessed several economic contractions during its history and there is no doubt that the economy was in recession, for example, during the 1990's and after the financial crisis in 2008 (for more discussion of these two severe recession periods, see Gulan, Haavio & Kilponen (2014)). There exists, however, no official recession and expansion chronology with exact start and end dates for Finland today. Therefore some turning point identification method needs to be applied. In this paper, we identify the turning points of Finnish economy by using the standard BB-algorithm for monthly data and its quarterly version, BBQ-algorithm, for quarterly GDP[18]. We use the trend indicator of output, coincident index calculated by Lanne & Nyberg (2009, 2015) and its slightly modified, stock returns augmented version to estimate the monthly chronology[19]. The estimated turning points are represented in table 2.

Table 2: Turning point candidates of Finnish economy in 1988-2019 based on BB- and BBQ-algorithms.

	GDP		**Lanne & Nyberg coincident index**		**Modified coincident index**		**Trend indicator of output**	
Sample	1988Q1-2018Q4		1988M1-2017M6		1989M2-2019M3		1995M1-2019M3	
	Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough
1	1990Q1	1993Q2	1990M4	1991M9	1990M5	1991M6		
2					2000M11	2001M8		
3	2007Q4	2009Q2	2008M4	2009M7	2008M5	2009M6	2007M12	2009M5
4	2011Q4	2013Q1	2011M11	2013M3	2012M2	2012M8	2010M12	2014M12
5	2013Q3	2015Q1	2013M12	2015M4	2014M8	2015M4		

The estimated peaks and troughs are quite consistent and in line with the economic history of Finland. However, there are some differences between the estimated turning points. First difference is related to period after the tech bubble in Finland at the beginning of 2000's. According to the stock returns augmented coincident index, there was an economic downturn after the tech bubble in early 2000's, but by other variables, this was not an actual recession period. Second difference is related to the financial crisis in late 2000's and the eurocrisis, a recession that followed it. Financial crisis and euro crisis are typically referred to as "double recession" in recent economic discussion in Finland, but according to both quarterly GDP and coincident indices, there has actually been three separate recession periods during that time. Only the trend indicator of output identifies the recession period of 2010's as a one prolonged economic downturn period.

Variables and indices seem also to date the length of 1990's recession differently. Lanne & Nyberg (2009) applied the BB-algorithm to the coincident index and concluded that the trough in 1990's was not reached until March 1993. However, when BB-algorithm is applied to the coincident index after data revisions, the trough is dated to be in September 1991. In comparison, the chronology

[18]We used the latest available vintage of the series in May 2019.

[19]Variables included in Lanne & Nyberg coincident index are GDP, employment, industrial production, export and import, whereas in the modified version the information of stock returns is also included. The modified index is available at https://econometrics.utu.fi/cei-fi/

based on quarterly GDP dates the trough at the second quarter of 1993. This highlights the fact that turning points can differ considerably depending on the variables they are estimated from. This also highlights the problematic relationship between the dating procedures and data revisions, since the algorithms may date the turning points differently depending which version of data is used. It is also worth of noting that unfortunately the trend indicator of output is not available before January 1995, so the information of it can not be used to date the end of 1990's recession.

Since our target is to apply a monthly probit model, we employ the monthly business cycle chronology instead of quarterly. Considering the economic history of Finland, the Lanne & Nyberg index seems to provide the most reliable alternative. However, we make a small modification to the estimated turning points by changing the trough of September 1991 to trough of March 1993. This is more in line with the development of GDP and unemployment rate as depicted in figure 1. The selected turning points are tabulated in table 3. Figure 2 represents the chosen recession periods together with coincident indices and the trend indicator of output.

Table 3: Selected turning points for Finland in 1988-2019.

	Peak	Trough
1	1990M4	1993M3
2	2008M4	2009M7
3	2011M11	2013M3
4	2013M12	2015M4

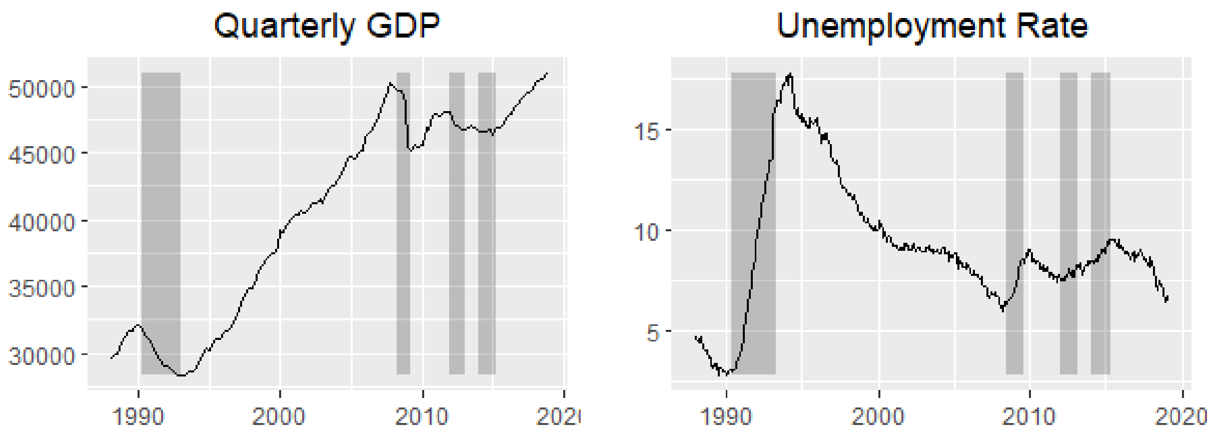

Figure 1: GDP and Unemployment Rate with Recession periods in 1988-2018. For GDP, the monthly periods are transformed to quarterly.

5.2 Data and Predictive variables

As discussed in the literature review, several possible leading indicators of recession periods have been suggested for the United States and other major economies. For Finland, the only probit analysis is by Pönkä Stenborg (2019), who applied the most common predictors, such as term spread and stock returns, for quarterly recession periods. In this article, we consider a larger set of variables and employ monthly data. The data sample starts from December 1988 and ends to

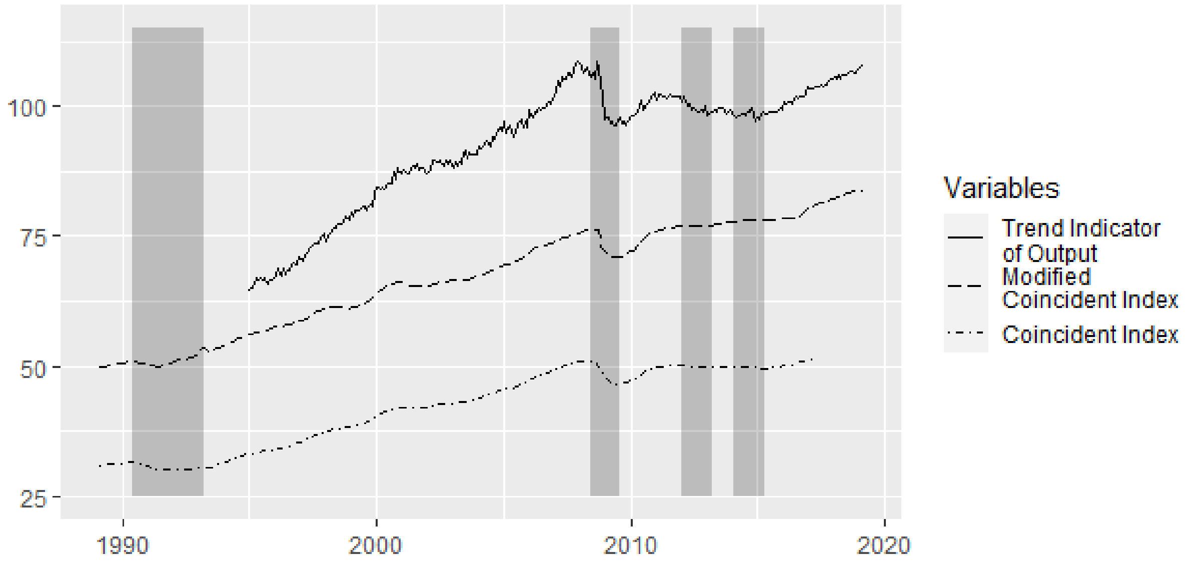

Figure 2: Coincident Indices and Trend Indicator of Output and Recession periods in 1988-2018.

March 2019, since we want to include the 1990's recession period in the sample[20]. Data set represents variables typically employed in probit model research. However, since we require our data to be both available from December 1988 and on a monthly frequency, some potential predictors are excluded.

The employed data set consists of 44 variables covering several financial variables, sentiment based variables and price indices. Among the financial variables are stock indices, dividend yields and the spread between the yield of Finnish government bond with maturity of 10 years and 3 month euribor rate[21]. Economic sentiment indices and confidence indicators of construction, consumers and industry are instead among the sentiment based variables[22]. In addition, the predictive power of the oil price, EUR/USD exchange rate and few building related variables, such as building cost index and building permits[23], are considered. Since Finland is a small open economy and thus heavily affected by the fluctuations in global economy, we include also foreign variables, such as German term spread, US stock returns and Euro Area business climate indicator, among the set of possible predictors. All the employed variables, their abbreviations and possible transformations to achieve stationarity are listed in Appendix C.

[20]We recognize that by choosing the start date of data sample as early as 1988M12, some potential predictors are excluded, since long enough series of them are not available.

[21]3 month Helibor Rate is used prior 1999.

[22]Data considering the consumer confidence in Finland was updated for this article in February 2020 after Statistics Finland published the level revised series. Series were revised due to the change in data collection method.

[23]Building variables are included in the data set, since for example building permits is one of the key variables in Leading Economic Index of Conference Board.

5.3 Results from Single-Predictor Models

In this section, we present the findings based on single-predictor models. Following the common representation, we first present the in-sample results and after that we consider the pseudo out-of-sample setting. In-sample results are based on models estimated over the entire sample, whereas in pseudo out-of-sample we apply the expanding window approach with estimation samples ranging from 1988M12–1999M12 to 1988M12–2018M9. Thus in pseudo out-of-sample setting we estimate the model by using data from 1988M12 to 1999M12 and, for example for the 6 month forecast horizon, calculate the first forecast for 2000M6. Then we estimate the model again by using data from 1988M12 to 2000M1 and calculate the second forecast for 2000M7. We continue this way until the last forecast for 2019M3 is calculated. Models are estimated by using the maximum likelihood estimation method and robust standard errors are obtained as in Kauppi & Saikkonen (2008)[24]. In this article, forecasts are obtained directly[25].

Previous research has suggested that the predictive ability of a leading variable may vary between different forecast horizons. For example, the term spread has predicted quite well the US recession periods with longer horizons (see, e.g Estrella & Mishkin, 1998 and Kauppi & Saikkonen, 2008). Thus we represent the findings for each single-predictor model by using the forecast horizons of 6, 9 and 12 months. We employ the lag orders k and l in a way that the most recent values of the explanatory variables are used. Thus in 6 month forecast horizon we set the $k = l = 6$, which has been a common choice in probit model research (see also Appendix B). Tables 4–6 summarize the selective in-sample results of single-predictor models, whereas selective out-of-sample results are represented in table 7. Variables in selective results are representatives of best-performing predictors as well as of the different predictor categories. The rest in-sample results are provided in Appendix D.

The in-sample findings from the single-predictor models implicate that term spreads, stock returns, dividend yields, economic sentiment indices and various confidence indicators are potentially useful monthly predictors of Finnish recession periods[26]. Economic sentiment index for Finland (ESI_FI), employment expectations in construction (CTRC_EE) and consumers expectation of general economic situation in Finland (CONS_GEN & CONS_GENO) seem particularly useful. The highest AUC-value is obtained by using the sixth lag of consumers expectations, whereas the highest adjusted pseudo-R^2 is obtained by using the sixth lag of the Finnish term spread (TSFI). As previous studies have concluded with US data, the term spread is able to predict the recession periods even with the longer forecast horizons. Similar findings also hold for the employment expectations in construction and consumers expectations, i.e they have predictive power even at the longer forecast horizons. It is interesting to note that the amount of building permits (FI_BP) and the price of oil (BR_OIL) are not successful predictors with forecast horizon of 6 months, but with the horizons of 9 and 12 months. Average dividend yields of both DAX (DAX_DY) and S&P500 (SP50_DY) indices are statistically significant predictors with all examined forecast horizons and perform reasonably well even with the longer horizons.

[24]Estimation is done via Matlab and the BFGS algorithm.

[25]This is due to the fact that iterative forecasts are computationally heavier and no proof of their superiority with Finnish data was found.

[26]We also tested the ability of aggregate indicators of consumer, construction and industry confidence of EU/EUR countries to predict Finnish recessions. However, the results were consistently weaker compared to those obtained by the Finnish indicators and thus these results are excluded from the article.

Table 4: In-sample results from single-predictor models.

Forecast horizon: 6 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
TSFI_10y_3m	-0.68***	0.28	0.28	148.30	142.42	0.78
TSGE_10y_3m	-0.66***	0.21	0.21	160.40	154.52	0.78
FI_BP	0.69	0.02	0.02	194.93	189.04	0.58
BR_OIL	4.85	0.02	0.02	194.95	189.07	0.58
OMX_HEL_LD	-4.07***	0.05	0.04	190.14	184.26	0.66
OMX_HEL_LD12	-1.79***	0.15	0.15	171.46	165.58	0.76
DAX_DY_M	0.73***	0.14	0.13	173.46	167.58	0.76
SP500_DY_M	1.01***	0.18	0.18	165.50	159.62	0.79
BCI_EUR_YOY	-0.31***	0.07	0.07	185.42	179.54	0.70
IFO_GER_YOY	-0.05**	0.05	0.05	189.53	183.65	0.65
ESI_FI_M	-0.05***	0.22	0.22	159.29	153.40	0.81
ESI_FI_YOY	-0.04***	0.17	0.17	167.51	161.63	0.78
CTRC_FI_D12	-0.03***	0.19	0.18	165.12	159.24	0.76
CTRC_EE_FI	-0.02***	0.22	0.21	159.40	153.52	0.80
CTRC_EC_FI_D12	-0.02***	0.13	0.13	175.08	169.19	0.70
IND_EE_FI_D12	-0.03***	0.14	0.14	173.31	167.43	0.74
IND_AO_FI	-0.02***	0.10	0.10	180.25	174.37	0.71
IND_PE_FI	-0.04***	0.16	0.16	168.92	163.04	0.78
CONS_GEN_FI	-0.02***	0.25	0.25	153.66	147.78	0.82
CONS_UE_FI	0.03***	0.14	0.14	172.92	167.04	0.75
CONS_GENO_FI_D12	-0.04***	0.27	0.27	149.47	143.59	0.83

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

Table 5: In-sample results from single-predictor models.

Forecast horizon: 9 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
TSFI_10y_3m	-0.59***	0.24	0.24	154.17	148.29	0.77
TSGE_10y_3m	-0.57***	0.17	0.17	166.78	160.91	0.77
FI_BP	1.20**	0.06	0.05	187.50	181.62	0.64
BR_OIL	8.49**	0.06	0.05	187.54	181.67	0.64
OMX_HEL_LD	-3.37***	0.03	0.03	191.97	186.10	0.64
OMX_HEL_LD12	-1.28***	0.08	0.08	182.89	177.02	0.71
DAX_DY_M	0.57***	0.09	0.09	181.60	175.73	0.73
SP500_DY_M	1.01***	0.14	0.14	172.42	166.54	0.75
BCI_EUR_YOY	-0.19*	0.03	0.03	192.63	186.76	0.65
IFO_GER_YOY	-0.02	0.01	0.01	195.82	189.95	0.58
ESI_FI_M	-0.03***	0.10	0.10	180.02	174.15	0.71
ESI_FI_YOY	-0.04***	0.15	0.15	171.07	165.20	0.77
CTRC_FI_D12	-0.02***	0.11	0.11	178.31	172.44	0.69
CTRC_EE_FI	-0.02***	0.12	0.12	176.67	170.80	0.72
CTRC_EC_FI_D12	-0.01**	0.06	0.06	186.64	180.77	0.62
IND_EE_FI_D12	-0.23***	0.09	0.08	182.22	176.35	0.71
IND_AO_FI	-0.01*	0.04	0.04	190.11	184.24	0.64
IND_PE_FI	-0.03***	0.07	0.07	184.51	178.63	0.70
CONS_GEN_FI	-0.06***	0.18	0.18	164.71	158.84	0.70
CONS_UE_FI	0.02**	0.05	0.05	188.57	182.70	0.67
CONS_GENO_FI_D12	-0.03***	0.19	0.19	163.50	157.62	0.78

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

Table 6: In-sample results from single-predictor models.

Forecast horizon: 12 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
TSFI_10y_3m	-0.47***	0.18	0.18	165.49	159.62	0.74
TSGE_10y_3m	-0.44***	0.11	0.11	177.30	171.44	0.73
FI_BP	1.72***	0.11	0.11	177.46	171.60	0.69
BR_OIL	12.18***	0.11	0.11	177.63	171.76	0.69
OMX_HEL_LD	-2.10**	0.01	0.01	194.69	188.82	0.60
OMX_HEL_LD12	-0.89**	0.04	0.04	189.40	183.54	0.65
DAX_DY_M	0.49**	0.07	0.07	184.82	178.95	0.71
SP500_DY_M	0.80***	0.12	0.12	175.33	169.47	0.73
BCI_EUR_YOY	-0.09	0.01	0	195.80	189.94	0.59
IFO_GER_YOY	0.001	0	0	196.94	191.07	0.49
ESI_FI_M	-0.02*	0.04	0.04	189.69	183.83	0.63
ESI_FI_YOY	-0.04***	0.12	0.12	174.85	168.99	0.75
CTRC_FI_D12	-0.01***	0.08	0.07	183.27	177.41	0.64
CTRC_EE_FI	-0.01**	0.06	0.06	186.27	180.40	0.64
CTRC_EC_FI_D12	-0.01	0.04	0.03	190.71	184.85	0.56
IND_EE_FI_D12	-0.01*	0.04	0.04	189.96	184.09	0.65
IND_AO_FI	-0.01	0.01	0.01	194.85	188.98	0.57
IND_PE_FI	-0.02	0.03	0.03	191.32	185.46	0.63
CONS_GEN_FI	-0.05***	0.16	0.16	167.77	161.91	0.78
CONS_UE_FI	0.01	0.03	0.02	192.27	186.41	0.62
CONS_GENO_FI_D12	-0.02***	0.10	0.10	178.40	172.53	0.72

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

The least favourable selected in-sample results were obtained with OMX Helsinki 1-month returns (OMX_HEL_LD), Euro Zone business climate indicator (BCI_EUR), IFO index of Germany (IFO_GER) and assessment of order book level in industry (IND_AO). Compared to the 1-month returns, the 12-month returns of OMX Helsinki (OMX_HEL_LD12) yields far more higher pseudo-R^2- and AUC-values while being statistically significant.

As an overall conclusion of the in-sample results with single predictor, the individual variables can predict Finnish recession periods generally well. The coefficients are also of the expected sign; the high term spread, stock returns and high economic sentiment index value are associated negatively with the risk of a recession, whereas the high value of consumers unemployment expectations indicator are associated positively with the risk of a recession. In general, most of the considered variables were able to predict the 1990's recession and financial crisis in 2008 while false negative alarms were given during the double recession period in 2010's.

As Estrella & Mishkin (1998) demonstrated, the variables with good in-sample fit do not necessarily have good out-of-sample predictive ability. Therefore we present also the selective out-of-sample estimation results for single-predictor models with forecast horizon of 6, 9 and 12 months in table 7. Several variables, such as consumers expectations of general economic situation in Finland and economic sentiment index, perform relatively well in 6 and 9 months forecast horizon, but for a few variables, the deterioration in performance is substantial. Building permits, price of oil, assessment of order book level in industry and construction confidence indicator (CTRC_FI) have very poor AUC-values for all examined forecast horizons.

Term spread of Finland yields moderate AUC-values also in pseudo out-of-sample estimation for all horizons. The deterioration is probably due to the fact that the time period, when predictions based on spread are substantially accurate, i.e the years 1989–1999, is excluded from the pseudo out-of-sample AUC-calculations. It seems that the ability of term spread of Finland to predict recessions has diminished after the financial crisis and the start of quantitative easing programme of European Central Bank. In particular, the low interest rate environment has probably affected the usefulness of term spread. It is also worth noting, that the first estimation sample includes unfortunately only one recession period (1990's recession), which may have affected our results a little. In other words, we estimate the first out-of-sample forecasts for June 2000 (6 month forecast horizon) and not, for example, for the period after financial crisis. We have chosen this approach in order to keep the forecasting period as long as possible, so that the out-of-sample predictive ability of different variables is tested at least with two different recession periods (financial crisis and euro crises).

Table 7: Out-of-sample results from single-predictor models.

	Forecast horizons		
	6 months	9 months	12 months
Variable	AUC	AUC	AUC
TSFI_10y_3m	0.69	0.68	0.62
TSGE_10y_3m	0.71	0.70	0.63
FI_BP	0.33	0.36	0.42
BR_OIL	0.69	0.36	0.42
OMX_HEL_LD	0.59	0.57	0.51
OMX_HEL_LD12	0.61	0.55	0.50
DAX_DY_M	0.80	0.77	0.73
SP500_DY_M	0.71	0.67	0.64
BCI_EUR_YOY	0.65	0.58	0.50
IFO_GER_YOY	0.63	0.52	0.44
ESI_FI_M	0.71	0.65	0.58
ESI_FI_YOY	0.78	0.74	0.66
CTRC_FI_D12	0.66	0.55	0.48
CTRC_EE_FI	0.75	0.66	0.59
CTRC_EC_FI_D12	0.60	0.49	0.42
IND_EE_FI_D12	0.70	0.65	0.56
IND_AO_FI	0.62	0.52	0.44
IND_PE_FI	0.71	0.62	0.54
CONS_GEN_FI	0.84	0.75	0.68
CONS_UE_FI	0.68	0.53	0.48
CONS_GENO_FI_D12	0.79	0.72	0.61

Notes: Since adjusted pseudo-R^2 can take negative values in out-of-sample estimation, only the AUC-value is represented as a goodness-of-fit measure. Abbreviations of variables can be found in Appendix C.

5.4 Results from Multi-Predictor Models

Next we present the selective in-sample and out-of-sample results from multi-predictor models, based on combinations of the variables used in the previous section[27]. We have chosen the combinations of variables in a following way. First, we want to include the term spread of Finland in every model to test its significance together with other possible predictors. Second, we want to include the best single predictor (CONS_GEN) based on its out-of-sample performance. Third, we are also interested whether construction or industry confidence bring additive predictive power over the spread and construction confidence[28]. Since financial variables have been typically considered to bring also additional predictive power, we add the 1-month and 12-month stock returns of OMX Helsinki in the variable set. We also extend our analysis by employing the dynamic probit model of the form (4). We set the lag orders k and l to match the forecast horizon and thus use the latest values of explanatory variables. Autoregressive probit model is not included in our analysis, since based on Pönkä & Stenborg's (2019) results, autoregressive component with Finnish data increases the in-sample AUC-values only very slightly.

Table 8: In-sample results from multi-predictor models.

	Forecast horizon: 6 months							
	Models							
	1	2	3	4	5	6	7	8
Variable	Coefficients							
Constant	-0.48**	-0.89***	-0.19	-0.48**	-0.73***	-1.12***	-1.13***	-1.13***
TSFI_10y_3m	-0.66***	-0.56***	-0.66***	-0.63***	-0.61***	-0.55***	-0.57***	-0.53***
CONS_GEN_FI	-0.07***	-0.07***	-0.06***	-0.06***	-0.07***	-0.07***	-0.07***	-0.06***
CTRC_EE_FI		-0.02***				-0.01	-0.01	-0.01
IND_PE_FI			-0.03***					
OMX_HEL_LD12				-0.43			0.56	
OMX_HEL_LD								-1.75
ESI_FI_M					-0.05***			
REC						1.18**	1.22**	1.20**
ps.R^2	0.43	0.53	0.47	0.44	0.52	0.58	0.58	0.58
adj.ps.R^2	0.43	0.52	0.47	0.43	0.51	0.57	0.57	0.57
BIC	122.52	108.20	117.78	124.58	109.53	101.43	103.40	103.66
logL	113.70	96.44	106.02	112.82	97.77	86.73	85.76	86.02
AUC	0.89	0.93	0.91	0.89	0.92	0.94	0.94	0.94

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

In-sample findings are presented in tables 8–10, whereas pseudo out-of-sample findings are in table 11. The results in tables 8–10 indicate that all selective multipredictor models perform better than any of the considered single-predictor models. The highest AUC-value and adjusted pseudo-R^2 value are obtained by using the 6-month forecast horizon, i.e the shortest forecast horizon considered in this article. In terms of all goodness-of-fit criteria, the models 6, 7 and 8 yield the highest results.

[27] All the possible combinations of variables are not calculated, since the purpose of this paper is to be illustrative and the needed computations would be quite heavy. We also recognize, that by proceeding this way there is a possibility, that the best-performing combination might not be presented.

[28] We tested all construction and industry confidence variables together with the spread and construction confidence, but we present only the results of the best combination of these three variables.

However, in model 7, the sign of the stock returns changes from negative to positive, which may indicate overfitting or multicollinearity, since it is difficult theoretically to justify the sign change. Even if the best performing models contain more variables, it is interesting to note that the model 2, based solely on two confidence variables and term spread, yields quite high in-sample AUC-values on all examined forecast horizons.

Table 9: In-sample results from multi-predictor models.

Forecast horizon: 9 months								
	Models							
	1	2	3	4	5	6	7	8
Variable	Coefficients							
Constant	-0.44**	-0.63***	-0.37	-0.44**	-0.52**	-0.72***	-0.78***	-0.73***
TSFI_10y_3m	-0.54***	-0.46***	-0.53***	-0.54***	-0.49***	-0.45***	-0.46***	-0.44***
CONS_GEN_FI	-0.05***	-0.05***	-0.04***	-0.05***	-0.05***	-0.04***	-0.05***	-0.04***
CTRC_EE_FI		-0.01*				-0.001	-0.01	-0.001
IND_PE_FI			-0.51					
OMX_HEL_LD12				0.02			0.53	
OMX_HEL_LD								-0.63
ESI_FI_M					-0.02			
REC						0.50	0.54	0.50
ps.R^2	0.34	0.36	0.34	0.34	0.35	0.37	0.38	0.37
adj.ps.R^2	0.34	0.36	0.33	0.33	0.34	0.37	0.37	0.37
BIC	139.93	138.26	142.27	142.87	140.49	139.20	140.94	142.00
logL	131.13	126.51	130.53	131.12	128.74	124.52	123.32	124.39
AUC	0.85	0.87	0.85	0.85	0.86	0.88	0.88	0.88

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

Table 10: In-sample results from multi-predictor models.

	Forecast horizon: 12 months							
				Models				
	1	2	3	4	5	6	7	8
Variable	Coefficients							
Constant	-0.51**	-0.58***	-0.54**	-0.52**	-0.51**	-0.60***	-0.67***	-0.59***
TSFI_10y_3m	-0.39***	-0.36***	-0.40***	-0.42***	-0.39***	-0.36***	-0.37***	-0.36***
CONS_GEN_FI	-0.04***	-0.04***	-0.04***	-0.05***	-0.04***	-0.04***	-0.05***	-0.04***
CTRC_EE_FI		-0.001				-0.001	-0.01	-0.001
IND_PE_FI			0.001					
OMX_HEL_LD12				0.35			0.62	
OMX_HEL_LD								0.59
ESI_FI_M					-0.001			
REC						0.07	0.13	0.07
ps.R^2	0.26	0.26	0.26	0.26	0.26	0.26	0.27	0.26
adj.ps.R^2	0.25	0.26	0.26	0.26	0.25	0.26	0.26	0.25
BIC	153.99	0.26	156.75	156.44	156.89	158.77	159.84	161.57
logL	145.19	0.26	145.03	144.49	145.17	144.12	142.25	143.98
AUC	0.81	0.82	0.81	0.81	0.81	0.82	0.83	0.82

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

In general, the models performing well in 6-month forecast horizon, yield also the highest goodness-of-fit criteria values in longer forecast horizons. Adding the dynamic component to the models clearly improves the AUC and adjusted pseudo-R^2 values in 6-month forecast horizon. The coefficient of the dynamic component is statistically significant at the 5% significance level in all models. The coefficient is also of the expected sign. The positive sign indicates that recession at present period increases the risk of a recession in the future. However, at the longer forecast horizons the sign remains positive but coefficients become statistically insignificant. This is in line with the results of Christiansen et al. (2014). Even if the dynamic model yields better AUC-values in 6 month forecast horizon, it is worth noting that the values of the binary indicator y_t are known within a delay. Thus in real-time forecasting the use of the dynamic probit model is problematic, especially if the delay is substantial as in the case of NBER.

Pseudo out-of-sample results confirm that multi-predictor models outperform the single-predictor models (tables 7 & 11). Highest AUC-values are obtained when dynamic extension is allowed. However, it is interesting to note that the model number 2, consisting only of three variables, performs again rather well especially in 6 month forecast horizon.

Table 11: Out-of-sample results from multi-predictor models.

	Forecast horizon:		
	6 months	9 months	12 months
Model number	AUC	AUC	AUC
1	0.83	0.77	0.71
2	0.90	0.81	0.73
3	0.88	0.80	0.74
4	0.83	0.77	0.72
5	0.90	0.81	0.73
6	0.92	0.83	0.74
7	0.92	0.84	0.76
8	0.92	0.82	0.74

Notes: Since adjusted pseudo-R^2 can take negative values in out-of-sample estimation, only the AUC-value is represented as a goodness-of-fit measure.

5.5 Recession Probabilities in 1989-2019

Figures 3–4 represent the recession probabilities based on probit model including either term spread (TSFI), consumers expectations (CONS_GEN_FI) or multiple predictors (model number 8 in table 8[29]) together with the recession periods between 1989M6-2019M3. Term spread turned to negative before the great recession in 1990's and before the financial crisis in 2008 causing the recession probabilities to increase timely. Instead, during the two euro crises in 2010's, the probabilities based solely on term spread probit model stay at a very moderate level. Probabilities between 20–25% are usually considered as a threshold of a weak recession signal, and the model containing only a spread does not give even a weak signal before euro crises. Term spread model obtains its highest value in 2010's in Octover 2015, six months *after* the last part of euro crisis actually ended and the Euro Area was recovering slowly.

Probit model based solely on consumers expectations of the general economic situation in Finland gives high probabilities before the first part of euro crisis in 2011 performing even better than the multi-predictor model. Both spread and consumers expectations based model give false positive recession signals during the tech bubble. It is worth noting that the model based on expectations gives weak signals nearly all the time between tech bubble and the financial crisis.

Multi-predictor model yields generally more accurate probabilities compared to single-predictor models. The fitted probabilities are low during the expansion periods apart from few exceptions like the tech bubble. The shortcoming of this multi-predictor model is that it does not induce high enough probabilities at the beginning of the euro crises, unlike the model based solely on consumers expectations. The fitted probability at the beginning of the first part of euro crises is only 8%. However, the fitted probabilities for the following months increase quickly and after three months they are already above 60%.

[29] Some other multi-predictor model could have been considered here, since the fitted probabilities with 6 month forecast horizon are quite identitical between models 6, 7 and 8.

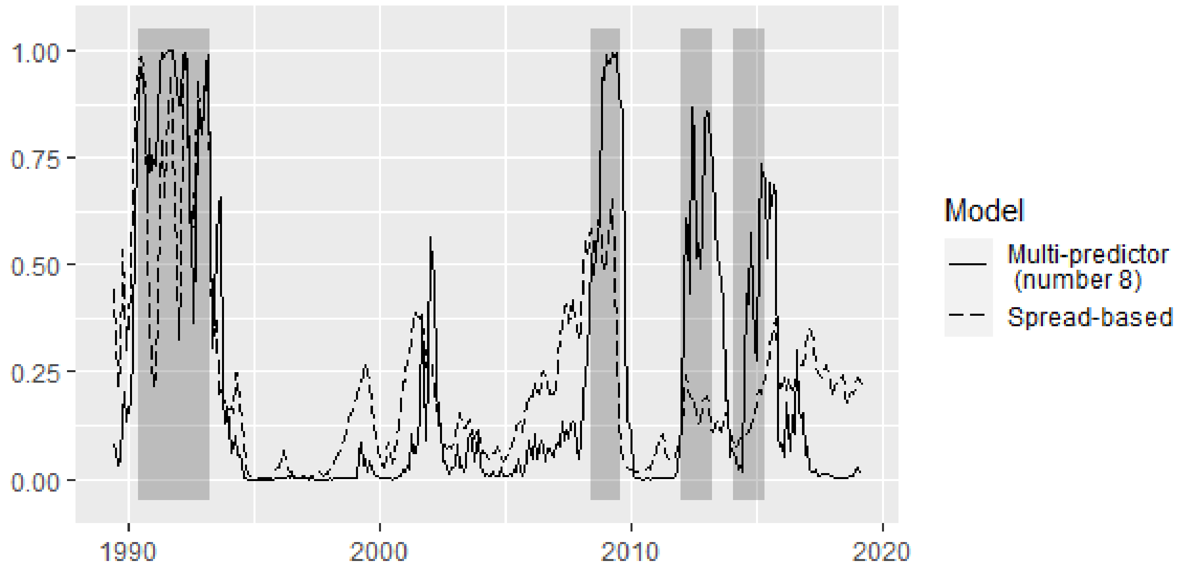

Figure 3: Recession probabilities (in-sample) based on term spread and model number 8 with 6-month forecast horizon.

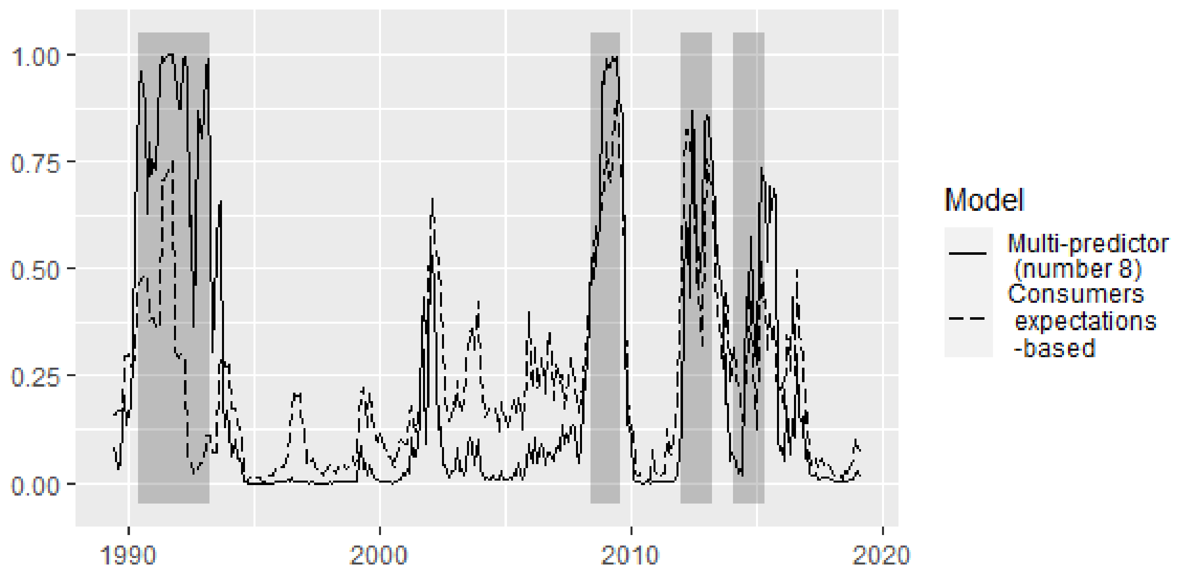

Figure 4: Recession probabilities (in-sample) based on consumers expectations and model number 8 with 6-month forecast horizon.

5.6 Employing shadow rates to probit models

In previous sections the analysis finds that a probit model with the term spread successfully gives high probabilities before the great recession and the financial crisis, but on the other hand, fails to predict the euro crises. The model also gives weak recession signals after the crises actually ended and Finland as well as the Euro Area were slowly recovering. One reason for this may be the quantitative easing programme started by European Central Bank and in particular the low interest rate environment. New monetary instruments such as large asset purchases have influenced the term spread and its components (see, for example, Wu & Xia, 2016; Kortela, 2016 and Kortela

& Nelimarkka, 2020 for more discussion). Hence, the original term spread in Euro Area countries may not have the same interpretation it had before the zero lower bound era, as discussed by Fendel et al. (2018).

To restore the predictive ability of term spread, Fendel et al. (2018) suggest to use the modified term spread instead of a basic spread. They estimate the modified term spread by replacing the 3-month euribor rate by a 3-month shadow rate of Wu & Xia (2016) from November 2013 onwards[30]. This modified term spread is showed to predict the Euro Area recessions defined by CEPR better than the original spread, since it can succesfully predict the start of euro crises.

In this article, we follow Fendel et. al (2018) by modifying the original term spread. We use the 3-month shadow rate estimated by Kortela (2016) and modify the original spread in two ways. First, we replace the 3-month euribor rate by a shadow rate from November 2013 onwards as Fendel et al. Second, we replace the euribor rate from May 2015 onwards, i.e from the first month when the 3-month euribor rate was actually below the zero. As in previous sections, we employ the data sample starting from December 1988 and ending to March 2019.

The in-sample findings from the single-predictor models are represented in table 12 whereas the out-of-sample results are presented in table 13. Both modified term spreads yield higher goodness-of-fit criteria values than the original spread. The increase in AUC-values is considerably notable for all examined forecast horizons. The second modified term spread produces better criteria values for 6 and 9 month forecast horizons, but slightly worse values in the longest horizon. The out-of-sample results confirm the in-sample findings.

Table 12: In-sample results from term spread models.

Forecast horizon: 6 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
TSFI_10y_3m	-0.68***	0.28	0.28	148.30	142.42	0.78
Modified spread 1	-0.66***	0.30	0.30	143.89	138.01	0.81
Modified spread 2	-0.73***	0.33	0.33	138.97	133.09	0.83
Forecast horizon: 9 months						
TSFI_10y_3m	-0.59***	0.24	0.24	154.17	148.29	0.77
Modified spread 1	-0.61***	0.29	0.28	146.22	140.35	0.82
Modified spread 2	-0.63***	0.29	0.29	145.85	139.98	0.82
Forecast horizon: 12 months						
TSFI_10y_3m	-0.47***	0.18	0.18	165.49	159.62	0.74
Modified spread 1	-0.51***	0.22	0.22	157.17	151.31	0.79
Modified spread 2	-0.51***	0.22	0.21	158.44	152.58	0.78

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels.

[30] Fendel et al. (2018) argue that from November 2013 ECB showed signs of downward rigidity of monetary policy by lowering the main refinancing rate to 0.25%.

Table 13: Out-of-sample results from term spread models.

	Forecast horizons		
	6 months	9 months	12 months
Variable	AUC	AUC	AUC
TSFI_10y_3m	0.69	0.68	0.62
Modified spread 1	0.77	0.78	0.73
Modified spread 2	0.79	0.78	0.71

Notes: Since adjusted pseudo-R^2 can take negative values in out-of-sample estimation, only the AUC-value is represented as a goodness-of-fit measure. AUC-values are calculated from the sample 2000M6-2019M3.

However, even the modified spreads cannot predict the start of the first part of euro crises in Finland, as depicted in figure 5. Thus the increase in all goodness-of-fit criterias is not due to the better ability of modified spread to predict the recession periods but instead the expansion periods. The estimated recession probabilities after euro crises are notably lower in the modified spread model than in the original model which can be considered as an advantage.

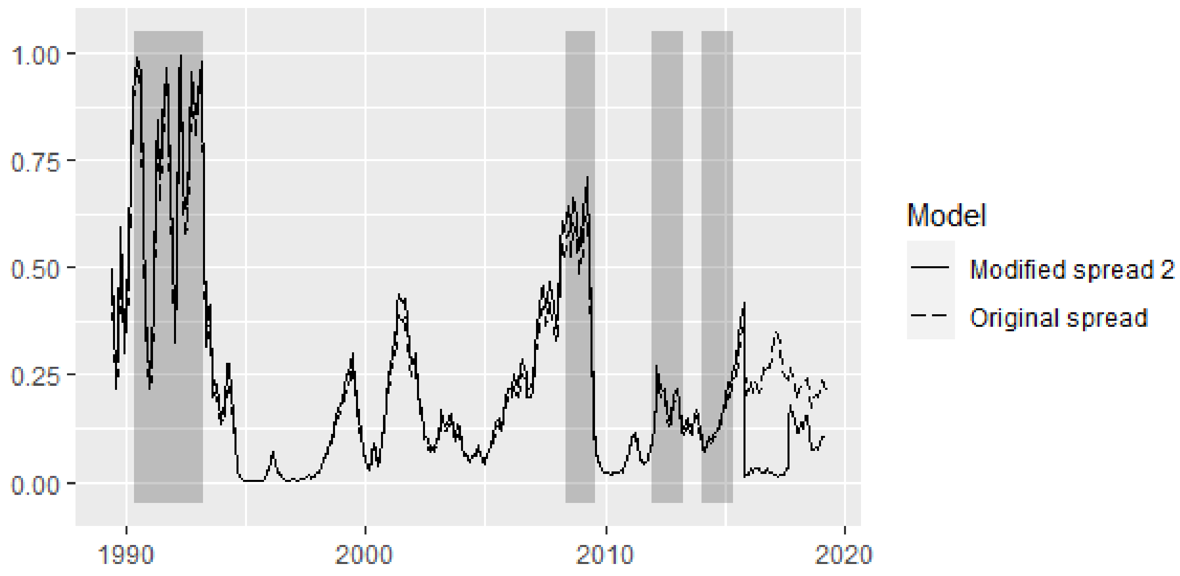

Figure 5: Recession probabilities (in-sample) based on term spreads with 6-month forecast horizon.

In table 14 are represented the in-sample results of multi-predictor model (model number 8) with the original term spread and the modified term spread. Out-of-sample results are tabulated in table 15. Findings based on in-sample and out-of-sample estimation indicate that the model including the revised spread produces more accurate forecasts compared the original model, though the differences are very subtle. It seems that multicollinearity or overfitting becomes problem in the longer forecast horizons, since the signs of recession indicator and stock returns change. In figure 6 are described the in-sample recession probabilities from multi-predictor models using the 6-month forecast horizon. Fitted probabilities differ only after the last recession period in a way that a model employing the modified term spread produces lower probabilities at the expansion era.

Table 14: In-sample results from multi-predictor models including shadow rates.

	Forecast horizon:					
	6 months		9 months		12 months	
Variable	Coefficients					
Constant	-1.13***	-1.04***	-0.73***	-0.61***	-0.59***	-0.47**
TSFI_10y_3m	-0.53***		-0.44***		-0.36***	
Modified spread 2		-0.54***		-0.47***		-0.41***
CONS_GEN_FI	-0.06***	-0.06***	-0.04***	-0.04**	-0.04***	-0.04***
CTRC_EE_FI	-0.01	-0.01	-0.001	-0.001	-0.001	-0.001
OMX_HEL_LD	-1.75	-1.78	-0.63	-0.60	0.59	0.64
REC	1.20**	1.14**	0.51	0.44	0.07	-0.01
ps.R^2	0.58	0.59	0.37	0.39	0.26	0.29
adj.ps.R^2	0.57	0.59	0.37	0.39	0.25	0.28
BIC	103.66	101.59	142	138.31	161.57	157.30
logL	86.02	83.95	124.39	120.69	143.98	139.71
AUC	0.94	0.94	0.88	0.89	0.82	0.83

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels. Abbreviations of variables can be found in Appendix C.

Table 15: Out-of-sample results from multi-predictor models.

	Forecast horizons		
	6 months	9 months	12 months
Model number	AUC	AUC	AUC
8, includes TSFI_10y_3m	0.92	0.82	0.74
8, includes modified spread 2	0.93	0.86	0.79

Notes: Since adjusted pseudo-R^2 can take negative values in out-of-sample estimation, only the AUC-value is represented as a goodness-of-fit measure. AUC-values are calculated from the sample 2000M6-2019M3.

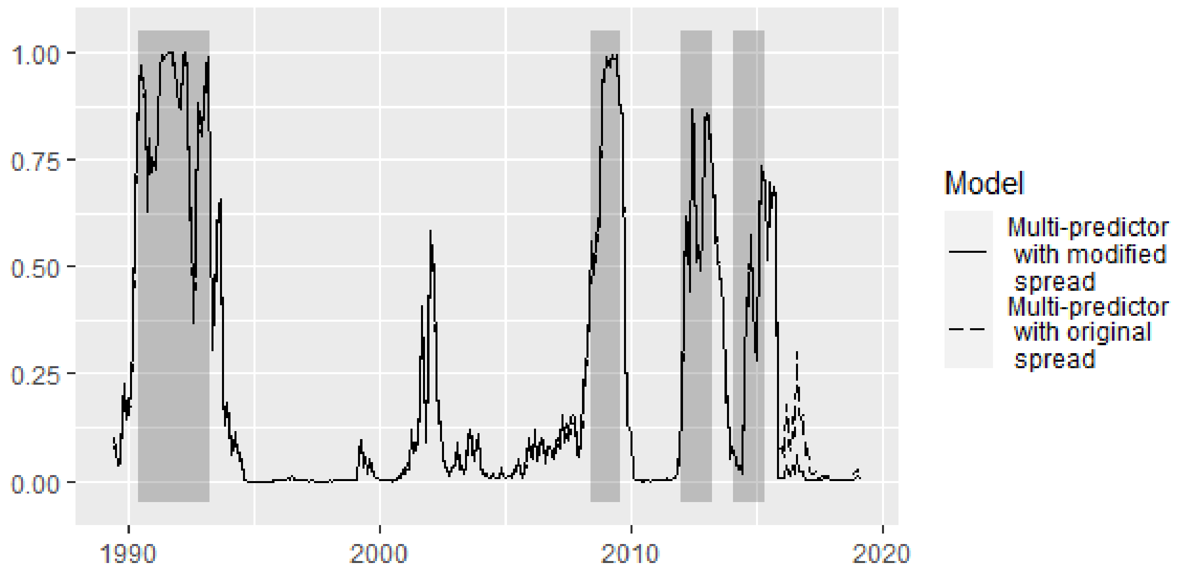

Figure 6: Recession probabilities (in-sample) based on multi-predictor models employing term spreads with 6-month forecast horizon.

6 Conclusion

This paper has reviewed the literature on recession forecasting with probit models. According to the review, which contains 25 articles and working papers, several financial and sentiment based variables are useful predictors of recession periods. Term spread seems to be a dominant predictor especially for the U.S recession periods, but its most accurate forecast horizon remains unclear. Several studies argue that the best results are obtained with the longer forecast horizons, typically 3 or 4 quarters ahead or 9 to 12 months ahead. However, the predictive power contained in term spread varies between countries. Business cycles in Japan, Netherlands and UK are especially difficult to predict with a model solely based on term spread. The predictive power of the spread seems to have also decreased in Euro Area countries after the euro crisis.

According to the review, various studies show that sentiment based variables, credit variables and other financial variables such as stock returns contain additional predictive power over the term spread. In addition, extended models outperform the basic static probit model. However, it remains unclear which extension yields the most accurate predictions, since the research results are mixed depending the data sample and employed variables.

One of the advantages of probit models is that they can be estimated by using monthly data. Typical predictors such as term spread, stock returns and confidence indicators are quickly available after the turn of the month, which is very useful, especially at the time of economic crisis. From this point, we have also presented an empirical illustration of probit models by predicting the monthly Finnish business cycle. We have established the Finnish chronology of expansion and recession periods and conclude, that there has been four separate recession periods in Finland in 1988–2019.

In this paper, we have applied the standard static probit model and its dynamic extension to forecast the Finnish business cycle states. Based on both in-sample and pseudo out-of-sample estimation results, various classical predictors, such as the term spread and confidence indicators, are useful in predicting recession periods of Finland. Our results are thus in general in line with the earlier studies, especially with those concerning Euro Area countries. The best single-predictor model uses data on consumers expectations of the general economic situation in Finland. Other sentiment based variables like the general economic sentiment index and employment expectations of construction perform also well. Dividend yields and 12-month stock returns of OMX Helsinki are the most useful financial variables. Term spread predicts correctly the first two recession periods in Finland, but does not capture the start of the euro crises unlike the model based on consumers expectations. Thus we conclude that the power of the term spread has decreased in Finland after the start of new monetary instruments, such as quantitative easing programme and forward guidance by the European Central Bank. This result is also in line with previous studies concerning Euro Area countries. Finally, we conclude that the multi-predictor models yield more accurate results and the dynamic extension outperforms the static model.

The results of this study can be extended in many ways. Data richer probit models, such as factor-augmented probit model, could be applied. It would be also interesting to investigate how recession probabilities are transmitted from other countries to Finland as Finland is a small open economy heavily influenced by the fluctuations in global economy. Since consumers expectations seem to have a large role in predicting Finnish business cycle, a Finnish economic policy uncertainty index could be constructed and test whether it has predictive power.

References

Ahking, Francis W. (2014) Measuring U.S. Business Cycles: A Comparison of Two Methods and Two Indicators of Economic Activitites. *Journal of Economic and Social Measurement.* Vol. 39, 199—216.

Ahrens, Ralf (2002) Predicting Recessions with Interest Rate Spreads: A Multicountry Regime Switching Analysis. *Journal of International Money and Finance.* Vol. 21, 519–537.

Akaike, Hirotugu (1974) A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control.* Vol. 19, 716—723.

Bellégo, C. – Ferrara, L. (2012) Macro-financial Linkages and Business Cycles: A Factor-Augmented Probit Approach. *Economic Modelling.* Vol. 29, 1792–1797.

Bengoechea, P. — Perez Quiros, G. (2004) A Useful Tool to Identify Recessions in the Euro Area. European Commission, Directorate-General for Economic and Financial Affairs Publications. Vol. 215.

Berge, T. – Jordà, Ò. (2011) Evaluating the Classification of Economic Activity into Recessions and Expansions. *American Economic Journal: Macroeconomics.* Vol. 3, 246-–277.

Bernard, H. — Gerlach, S. (1998) Does the term Structure Predict Recessions ? The International Evidence. *International Journal of Finance and Economics.* Vol. 3, 195-–215.

Bruno, G. — Otranto, E. (2008) Models to Date the Business Cycle: The Italian Case. *Economic Modelling.* Vol. 25, 899–9-11.

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

Burns, A. — Mitchell, W. (1946) Measuring Business Cycles. NBER Books, National Bureau of Economic Research.

Camacho, M. – Perez-Quiros, G. (2002) This is What the Leading Indicators Lead. *Journal of Applied Econometrics.* Vol. 17, 61–80.

Chauvet, Marcelle (1998) An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. *International Economic Review.* Vol. 39, 969-–996.

Chauvet, M. — Potter, S. (2005) Forecasting Recessions Using the Yield Curve. *Journal of Forecasting.* Vol. 24, 77—103.

Chauvet, M. — Piger, J. (2008) A Comparison of the Real-Time Performance of Business Cycle Dating Methods. *Journal of Business & Economics Statistics.* Vol. 26, 42—49.

Chen, Z. – Iqbal, A. – Lai, H. (2011) Forecasting the probability of US Recessions: A Probit and Dynamic factor Modelling Approach. *The Canadian Journal of Economics.* Vol. 44, 651–672.

Christiansen, C. – Eriksen, J. – Moller, S. (2014) Forecasting US Recessions: The Role of Sentiment. *Journal of Banking and Finance.* Vol. 49, 459—468.

Clements, M. – Krolzig, H-M. (1998) A Comparison of the Forecast Performance of Markov Switching and Threshold Autoregressive Models of US GNP. *The Econometrics Journal.* Vol. 1, C47—C75.

de Jong, R.M. — Woutersen, T. (2011) Dynamic Time Series Binary Choice. *Econometric Theory.* Vol. 27, 673-–702.

Diebold, F.X. — Rudebusch, G.D. (1989) Scoring the Leading Indicators. *The Journal of Business.* Vol. 62, 369—391.

Duarte, A. – Venetis, I. – Paya, I. (2005) Predicting Real Growth and the Probability of Recession in the Euro Area Using the Yield Spread. *International Journal of Forecasting.* Vol. 21, 261–277.

Dueker, Michael J. (1997) Strengthening the Case for the Yield Curve as a Predictor of U.S Recessions. *Federal Reserve Bank of Saint Louis Review.* Vol. 79, 41—51.

Estrella, Arturo (1998) A New Measure of Fit for Equations with Dichotomous Dependent Variables. *The Journal of Business and Economic Statistics.* Vol. 16, 198-–205.

Estrella, A. — Hardouvelis, G. (1991) The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance.* Vol. 46, 555-–576.

Estrella, A. – Mishkin, F. (1998) Predicting U.S. Recessions: Financial Variables as Leading Indicators. *The Review of Economics and Statistics.* Vol. 80, 45-–61.

Estrella, A. — Rodrigues, A. – Schich, S. (2003) How Stable is the Predictive Power of the Yield Curve ? Evidence from Germany and the United States. *The Review of Economics and Statistics.* Vol. 85, 629-–644.

Fendel, R. – Mai, N. – Mohr, O. (2018) Recession Probabilities for the Euro Zone at the Zero Lower Bound: Challenges to the Term Spread and Rise of Alternatives. Otto Beisheim School of Management Working Paper 4(2018.

Fornari, F. – Lemeke, W. (2010) Predicting Recession Probabilities with Financial Variables over Multiple Horizons. ECB Working Paper no. 1255.

Fornaro, Paolo (2016) Forecasting U.S Recessions with a Large Set of Predictors. *Journal of Forecasting.* Vol. 35, 477–4921.

Fossati, Sebastian (2015) Forecasting U.S Recessions with Macro Factors. University of Alberta Working Papers.

Geweke, J. — Whiteman, C. (2006) Bayesian Forecasting. Chapter 1 in *Handbook of Economic Forecasting*, edited by Elliot, G. — Granger, C.W.J. – Timmermann, A. North Holland, Amsterdam.

Gulan, A. – Haavio, M. – Kilponen, J. (2014) Kiss Me Deadly: From Finnish Great Depression to Great Recession. Bank of Finland Research Discussion Paper 24/2014.

Hamilton, James D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and Business Cycle. *Econometrica.* Vol. 57, 357—384.

Hamilton, James D. (2003) Comment on "A Comparison of Two Business Cycle Dating Methods". *Journal of Economic Dynamics and Control.* Vol. 27, 1691-–1693.

Hamilton, James D. (2011) Calling Recessions in Real Time. *International Journal of Forecasting.* Vol. 27, 1006—1026.

Hansen, Bruce (1992) The Likelihood Ratio Test Under Nonstadard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics.* Vol. 7, S61—S82.

Harding, D. — Pagan, A. (2002) Dissecting the Cycle: A Methodological Investigation. *Journal of Monetary Economics.* Vol. 49, 365—381.

Harding, D. – Pagan, A. (2003a) A Comparison of Two Business Cycle Dating Methods. *Journal of Economic Dynamics and Control.* Vol. 27, 1681—1690.

Harding, D. — Pagan, A. (2003b) Rejoinder to James Hamilton. *Journal of Economic Dynamics and Control.* Vol. 27, 1695—1698.

Harding, D. — Pagan, A. (2006) Synchronization of Cycles. *Journal of Econometrics.* Vol. 132, 59—79.

Karnizova, L. – Li, J. (2014) Economic Policy Uncertainty, Financial Markets and Probability of US Recessions. *Economic Letters.* Vol. 125, 261–265.

Kauppi, H. — Saikkonen, P. (2008) Predicting U.S. Recessions with Dynamic Binary Response Models. *The Review of Economics and Statistics.* Vol. 90, 777—791.

Kortela, Tomi (2016) A Shadow Rate Model with Time-Varying Lower Bound of Interest Rates. Bank of Finland Research Discussion Paper 19/2016.

Kortela, T. – Nelimarkka, J. (2020) The Effects of Conventional and Unconvential Monetary Policy: Identification Through the Yield Curve. Bank of Finland Research Discussion Paper 3/2020.

Lanne, M. – Nyberg, H. (2009) Suomen kansantalouden suhdanneindeksi. *Kansantaloudellinen aikakauskirja.* Vol. 105, 422-432.

Lanne, M. – Nyberg, H. (2009) Suomen kansantalouden suhdanneindeksi 2009–2014. *Kansantaloudellinen aikakauskirja.* Vol. 111, 6–15.

Maddala, Gangadharrao S.(1983) *Limited Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, New York.

Moneta, Fabio (2003) Does the Yield Spread Predict Recessions in the Euro Area ? ECB Working

Paper no. 294.

Morley, J. — Piger, J. (2005) The Importance of Nonlinearity in Reproducing Business Cycle Features. Working Paper 2004–032B, Federal Reserve Bank of Saint Louis.

Mönch, E. — Uhlig, H. (2005) Towards a Monthly Business Cycle Chronology for the Euro Area. *Journal of Business Cycle Measurement and Analysis*. Vol. 2, 43-–69.

National Bureau of Economic Research (2008) Announcement of December 2007 Business Cycle Peak/Beginning of Last Recession. <https://www.nber.org/cycles/dec2008.html>. [page visited 9.6.2019]

Ng, Eric (2012) Forecasting US Recessions with Various Risk Factors and Dynamic Probit Models. *Journal of Macroeconomics.* Vol. 34, 112–125.

Nyberg, Henri (2010) Dynamic Probit Models and Financial Variables in Recession Forecasting. *Journal of Forecasting.* Vol. 29, 215-–230.

Poirier, D.J. — Ruud, P.A (1988) Probit with Dependent Observations. *The Review of Economic Studies*. Vol. 55, 595-–614.

Pönkä, Harri (2017) The Role of Credit in Predicting U.S. Recessions. *Journal of Forecasting.* Vol. 29, 215–247.

Pönkä, H. — Stenborg, M. (2019) Forecasting the State of the Finnish Business Cycle. Publications of Ministry of Finance, 2019:13, 1—16.

Rosenberg, J. – Maurer, S. (2008) Signal or Noise ? Implications of the Term Premium for Recession Forecasting. Federal Reserve Bank of New York Economic Policy Review. Vol. 14.

Schwarz, Gideon (1978) Estimating the Dimension of a Model. *Annals of Statistics.* Vol. 6, 461—464.

Schirwitz, Beate (2009) A Comprehensive German Business Cycle Chronology. *Empirical Economics*. Vol. 37, 287—301.

Stock, J.H. — Watson, M.W. (1989) New Indexes of Coincident and Leading Economic Indicators. NBER Macroeconomics Annual. Vol. 4, 351—409. MIT Press.

Stock, J.H. — Watson, M.W. (2003) Forecasting Output and Inflation: The Role of Asset Prices. *Journal of Economic Literature.* Vol. 41, 788—829.

Stock, J.H. — Watson, M.W. (2014) Estimating Turning Points Using Large Data Sets. *Journal of Econometrics*. Vol. 178, 368—381.

Wheelock, D. — Wohar, M. (2009) Can the Term Spread Predict Output Growth and Recessions ? A Survey of the Literature. *Federal Reserve Bank of Saint Louis Review*. Vol. 91, 419—440.

Wright, Jonathan (2006) The Yield Curve and Predicting Recessions. Finance and Economics Discussion Series. Federal Reserve Board, Washington, D.C.

Wu, C. – Xia, F. (201) Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money, Credit and Banking.* Vol. 48, 253–291.

Appendix A Maximum Likelihood Estimation of Probit Models

According to the properties of Bernoulli distributed random variables, y_t follows a conditional Bernoulli distribution, indicating that

$$p_t = P_{t-1}(y_t = 1) = \Phi(\pi_t)$$
$$1 - p_t = P_{t-1}(y_t = 0) = 1 - \Phi(\pi_t),$$

where P signifies the conditional probability given the information set Ω_{t-1}[31]. Thus, the conditional probability density function for observation y_t is

$$f_{t-1}(y_t) = f(y_t|\Omega_{t-1}) = (\Phi(\pi_t))^{y_t}(1 - \Phi(\pi_t))^{1-y_t}.$$

Let us assume that we have observed the sample denoted by $y = (y_1, \ldots, y_T)$ and that the initial value y_0 is also available. Then the joint conditional probability density function is

$$f_{t-1}(y) = f(y_t, \ldots, y_T|\Omega_{t-1}) = \prod_{t=1}^{T} f(y_t|\Omega_{t-1}).$$

Let us define the parameter vector $\theta = [\omega\, \beta\, \alpha_1\, \delta_1]$. Then the log likelihood function, conditional on initial values, has the form

$$l(\theta) = \sum_{t=1}^{T} l_t(\theta) = \sum_{t=1}^{T} [y_t \log(\Phi(\pi_t(\theta))) + (1 - y_t) \log(\Phi(\pi_t(\theta)))],$$

where $\pi_t(\theta)$ is given by (5) or some of its restricted versions. When equation (5) or (6) is used as $\pi_t(\theta)$, a choice for the initial value π_0 is needed. Kauppi & Saikkonen (2008) suggested that this initial value is chosen by a formula that can be interpreted as an estimate of the unconditional mean of π_t. For example, in the case of the model (5), they set $\pi_0 = (\omega + \delta_1 \bar{y} + \bar{x}'_{t-k}\beta)/(1 - \alpha_1)$.

The maximization of the log likelihood function is a nonlinear optimization problem that can be carried out by using standard numerical methods, such as the BFGS algorithm. The maximum likelihood estimate $\hat{\theta}$ solves the first order condition $s(\hat{\theta}) = 0$, where $s(\theta)$ is a score function

$$s(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \sum_{t=1}^{T} \left(y_t \frac{1}{\Phi(\pi_t)} \frac{\partial \Phi(\pi_t)}{\partial \theta} + (y_t - 1) \frac{1}{(1 - \Phi(\pi_t))} \frac{\partial \Phi(\pi_t)}{\partial \theta} \right).$$

[31] The information set depends on the model spesification. For example, in the case of a model (6), the information set is $\Omega_{t-1} = \{(\pi_t, x_{t-k}, k \geq 1)\}$.

Appendix B Forecasting Procedures of Probit Models

Kauppi & Saikkonen (2008) have proposed two methods, "direct" and "iterative", to calculate multiperiod forecasts of probit models. Here a brief review is given to cover the basic idea of these procedures.

In a mean square sense, an optimal h-period forecast of y_t based on information at a time $t-h$ is given by the conditional expectation $E_{t-h}(y_t) = P_{t-h}(y_t = 1)$. By the law of iterated conditional expectations $E_{t-h}(y_t) = E_{t-h}(\Phi(\pi_t))$, where π_t is given by the probit model spesification (i.e the equations (3)-(6)).

Let us consider, for simplicity, the dynamic probit model. In a mean square sense, the optimal one-period forecast at time $t-1$ is obtained directly from an equation

$$p_t = P_{t-1}(y_t = 1) = \Phi(\omega + x'_{t-1}\beta + \delta_1 y_{t-1}),$$

where P signifies again the conditional probability. The multiperiod forecasts can be constructed either directly or in an iterative way. In a direct approach, an optimal h-period ahead forecast at time $t-h$ is obtained by an equation

$$p_t = P_{t-h}(y_t = 1) = \Phi(\omega + x'_{t-k}\beta + \delta_1 y_{t-l}), \tag{9}$$

assuming that $k, l \geq h$ so that the lagged values are known at time $t-h$ [32]. It has been common to choose the lag orders k and l in a way, that the most recent values of the explanatory values and the recession indicator are employed. For example, Dueker (1997) applied the above model with selection $k = l = h$ and Estrella & Mishkin (1998) applied the static probit models with a choice $k = h$. However, Kauppi & Saikkonen (2008) have argued that the most recent values may not be the most optimal and instead, the lag orders should be selected by statistical procedures.

In direct approach, the optimal h-period ahead forecasts can be obtained for one value of h at a time. In iterative approach, the lag order of the recession indicator does not need to be related to forecast horizon. Let us consider, for simplicity, again the dynamic probit model and the case $h = 2$. By using the one-period model iteratively and assuming $k \geq 2$ one obtains

$$E_{t-2}(y_t) = \sum_{y_{t-1} \in 0,1} P_{t-2}(y_{t-1})\Phi(\omega + x'_{t-k}\beta + \delta_1 y_{t-1}),$$

where the conditional probability P is given by

$$P_{t-2}(y_{t-1}) = \Phi(\omega + x'_{t-k-1}\beta + \delta_1 y_{t-2})^{y_{t-1}} \times [1 - \Phi(\omega + x'_{t-k-1}\beta + \delta_1 y_{t-2})]^{1-y_{t-1}}.$$

Thus the iterative forecast takes into account the two possible paths through which the economy can enter the recession in two period's time. Either the economy is in a recession at time $t-1$ or it is not. Naturally, when forecast horizon $h > 2$, the amount of possible paths is larger and the explicit formulas become more complicated.

[32] In practise, when recessions are predicted, one must note that that the values of some explanatory variables as well as the value of the binary indicator y_t are known with a delay.

Appendix C List of Variables

The following table lists the used variables, their abbreviations and possible transformations.

Table 16: Considered predictive variables.

Variable	**Abbreviation**	**Transformation**
Binary Indicator, 1 if economy is in recession at time t and 0 if not	REC	no transformation
Interest rates and spreads		
Term spread between the yield of Finnish government bond with maturity of 10 years and 3 month euribor rate	TSFI_10y_3m	no transformation
Term spread for Germany	TSGE_10y_3m	no transformation
The yield of Finnish government bond with maturity of 10 years	FI_10y_D	difference
3 month Euribor rate (Helibor Rate until 1999)	FI_3m_D	difference
Prices		
Consumer Price Index of Finland (1990=100)	FI_CPI	log
Production Price Index of Finland (1949=100)	FI_PPI	log
The price of euro in US dollars (prior 1999 markka)	EURUSD_D	difference
Brent Crude Oil Price	BR_OIL	log
Building-based		
Building Cost Index of Finland (1980=100)	FI_BCI	log
Building Permits (housing permits) in Finland (1000m3)	FI_BP	log
Financial		
OMX Helsinki Price Index, 1-month returns	OMX_HEL_LD	log-difference
OMX Helsinki Price Index, 12-month returns	OMX_HEL_LD12	log-difference, 12 lags
DAX Price Index	DAX_LD	log-difference
S&P500 Price Index	SP500_LD	log-difference
Average dividend yield in DAX	DAX_DY_M	demean
Average dividend yield in S&P500	SP500_DY_M	demean
Sentiment-based		
Euro Zone Business Climate Indicator	BCI_EUR	no transformation
Euro Zone Business Climate Indicator, difference	BCI_EUR_D12	difference, 12 lags
IFO Business Climate Index, Germany, 2015 = 100, YoY	IFO_GER_YOY	Year-Over-Year Change
Economic Sentiment Index, EU countries	ESI_EU_M	demean
Economic Sentiment Index, Euro countries	ESI_EUR_M	demean
Economic Sentiment Index, Finland	ESI_FI_M	demean
Economic Sentiment Index, EU countries, YoY	ESI_EU_YOY	Year-Over-Year Change
Economic Sentiment Index, Euro countries, YoY	ESI_EUR_YOY	Year-Over-Year Change
Economic Sentiment Index, Finland, YoY	ESI_FI_YOY	Year-Over-Year Change
Construction Confidence (CC) Indicator, Finland	CRTC_FI	no transformation
Construction Confidence (CC) Indicator, Finland, difference	CRTC_FI_D12	difference, 12 lags
Construction confidence: Employment Expectations over the next 3 months, Finland	CRTC_EE_FI	no transformation
Construction Confidence: Employment Expectations over the next 3 months, Finland, difference	CRTC_EE_FI_D12	difference, 12 lags
Construction Confidence: Evolution of the current overall order books, Finland	CRTC_EC_FI	no transformation
Construction Confidence: Evolution of the current overall order books, Finland, difference	CRTC_EC_FI_D12	difference, 12 lags

Considered predictive variables (Table 16 continues).

Variable	**Abbreviation**	**Transformation**
Industry Confidence: Employment Expectations over the next three months, Finland	IND_EE_FI	no transformation
Industry Confidence: Employment Expectations over the next three months, Finland, difference	IND_EE_FI_D12	difference, 12 lags
Industry Confidence: Assessment of order-book level, Finland	IND_AO_FI	no transformation
Industry Confidence: Assessment of order-book level, Finland, difference	IND_AO_FI_D12	difference, 12 lags
Industry Confidence: Production Expectations over the next 3 months, Finland	IND_PE_FI	no transformation
Industry Confidence: Production Expectations over the next 3 months, Finland, difference	IND_PE_FI_D12	difference, 12 lags
Consumer Confidence: General Fconomic Situation over the next 12 months, Finland	CONS_GEN_FI	no transformation
Consumer Confidence: General Economic Situation over the next 12 months, Finland, difference	CONS_GEN_FI_D12	difference, 12 lags
Consumer Confidence: Unemployment expectations over the next 12 months, Finland	CONS_UE_FI	no transformation
Consumer Confidence: Unemployment expectations over the next 12 months, Finland, difference	CONS_UE_FI_D12	difference, 12 lags
Consumer Confidence: General Economic Situation over the last 12 months, Finland	CONS_GENO_FI	no transformation
Consumer Confidence: General Economic Situation over the last 12 months, Finland, difference	CONS_GENO_FI_D12	difference, 12 lags

Data sources: Reuters (interest rates and spreads), Statistics Finland (consumer and production price indices), Deutsche Börse (oil price, euro-dollar price, DAX price index & dividend yield), Bank of Finland calculations (building variables), NASDAQ OMX Helsinki (OMX Helsinki price index), Standards & Poor (S&P500 price index & dividend yield), Eurostat (sentiment variables excluding the IFO index), Ifo Institute for Economic Research (IFO index).

Appendix D In-Sample Results

Table 17: The rest in-sample results from single-predictor models.

Forecast horizon: 6 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
FI_10y_D	0.11	0	0	198.46	192.58	0.50
FI_3m_D	0.06	0	0	198.46	192.58	0.47
FI_CPI	-0.17	0	0	198.47	192.59	0.53
FI_PPI	0.18	0	0	198.49	192.61	0.52
FI_BCI	0.53	0	0	197.70	191.82	0.49
EURUSD_D	-2.08	0	0	197.97	192.09	0.52
DAX_LD	-2.46*	0.01	0.01	196.44	190.56	0.57
SP500_LD	-3.53*	0.01	0.01	196.50	190.62	0.56
BCI_EUR	-0.08	0	0	197.82	191.94	0.58
ESI_EU_M	-0.02*	0.03	0.03	193.30	187.42	0.66
ESI_EUR_M	-0.02*	0.03	0.03	193.50	187.61	0.65
ESI_EU_YOY	-0.03***	0.09	0.08	183.19	177.31	0.71
ESI_EUR_YOY	-0.03***	0.09	0.08	183.15	177.27	0.71
CTRC_FI	-0.01***	0.10	0.09	181.44	175.56	0.71
CTRC_EE_FI_D12	-0.02***	0.16	0.15	170.43	164.55	0.75
CTRC_EC_FI	-0.01	0.02	0.02	194.63	188.75	0.58
IND_EE_FI	-0.02**	0.07	0.06	186.59	180.71	0.62
IND_AO_FI_D12	-0.01***	0.08	0.08	183.47	177.59	0.72
IND_PE_FI_D12	-0.01**	0.06	0.06	186.91	181.03	0.66
CONS_GEN_FI_D12	-0.02	0.03	0.03	193.55	187.67	0.62
CONS_UE_FI_D12	0.02**	0.09	0.09	181.76	175.88	0.71
CONS_GENO_FI	-0.03***	0.26	0.26	150.69	144.81	0.84

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels.

Table 18: The rest in-sample results from single-predictor models.

Forecast horizon: 9 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
FI_10y_D	0.07	0	0	197.71	191.84	0.51
FI_3m_D	0.24	0.01	0	196.71	190.83	0.55
FI_CPI	-0.36	0	0	197.46	191.59	0.54
FI_PPI	0.17	0	0	197.70	191.83	0.52
FI_BCI	0.48	0	0	197.05	191.17	0.49
EURUSD_D	-0.05	0	0	197.74	191.86	0.49
DAX_LD	-1.90	0.01	0	196.50	190.63	0.55
SP500_LD	-2.88	0.01	0	196.41	190.54	0.55
BCI_EUR	0.11	0.01	0	196.57	190.70	0.52
ESI_EU_M	-0.001	0	0	197.54	191.67	0.57
ESI_EUR_M	-0.001	0	0	197.58	191.71	0.56
ESI_EU_YOY	-0.02^{**}	0.05	0.04	189.65	183.78	0.66
ESI_EUR_YOY	-0.02^{**}	0.04	0.04	189.85	183.98	0.66
CTRC_FI	-0.01^{*}	0.03	0.03	192.03	186.15	0.61
CTRC_EE_FI_D12	-0.02^{***}	0.11	0.11	177.60	171.72	0.72
CTRC_EC_FI	-0.001	0	0	197.66	191.78	0.48
IND_EE_FI	-0.01	0.02	0.01	194.70	188.82	0.53
IND_AO_FI_D12	-0.01^{***}	0.07	0.07	185.09	179.22	0.72
IND_PE_FI_D12	-0.02^{**}	0.04	0.04	189.96	184.09	0.66
CONS_GEN_FI_D12	-0.02	0.03	0.03	192.40	186.53	0.63
CONS_UE_FI_D12	0.02^{*}	0.04	0.04	190.00	184.13	0.65
CONS_GENO_FI	-0.02^{***}	0.14	0.13	173.34	167.47	0.76

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels.

Table 19: The rest in-sample results from single-predictor models.

Forecast horizon: 12 months						
Variable	Coefficient	ps.R^2	adj.ps.R^2	BIC	logL	AUC
FI_10y_D	0.32	0	0	196.30	190.43	0.56
FI_3m_D	0.19	0.01	0	196.30	190.44	0.57
FI_CPI	-0.60	0	0	196.19	190.33	0.55
FI_PPI	0.04	0	0	196.94	191.07	0.51
FI_BCI	0.39	0	0	196.50	190.64	0.48
EURUSD_D	1.48	0	0	196.68	190.81	0.54
DAX_LD	-0.41	0.01	0	196.88	191.02	0.51
SP500_LD	0.20	0	0	196.93	191.07	0.50
BCI_EUR	0.27^{*}	0.03	0.03	190.84	184.98	0.61
ESI_EU_M	0.01	0.01	0	196.04	190.17	0.52
ESI_EUR_M	0.01	0.01	0	195.93	190.06	0.52
ESI_EU_YOY	-0.02	0.02	0.02	193.22	187.36	0.61
ESI_EUR_YOY	-0.02	0.02	0.02	193.41	187.55	0.60
CTRC_FI	-0.001	0.01	0	195.71	189.84	0.53
CTRC_EE_FI_D12	-0.02^{***}	0.10	0.09	179.84	173.97	0.70
CTRC_EC_FI	0.001	0	0	196.16	190.30	0.57 T
IND_EE_FI	-0.001	0.02	0.01	196.88	191.01	0.46
IND_AO_FI_D12	-0.01^{**}	0.05	0.04	188.64	182.77	0.69
IND_PE_FI_D12	-0.01^{*}	0.03	0.02	192.05	186.19	0.62
CONS_GEN_FI_D12	-0.02	0.03	0.03	190.84	184.97	0.62
CONS_UE_FI_D12	0.01	0.02	0.02	193.11	187.24	0.60
CONS_GENO_FI	-0.02^{***}	0.07	0.06	185.03	179.16	0.69

Notes: In the table *,** and *** denote the statistical significance of the estimated coefficients using robust standard errors at 10%, 5% and 1% significance levels.