Forecasting U.S. recessions using dynamic probit models with a large number of predictors

Shuting Chen

School of Economics

University of Sydney

Thesis submitted in partial fulfilment of the award course requirements of the Bachelor of Economics (Honours)

Supervised by Dr. Peter Exterkate

Abstract

This thesis introduces new dynamic autoregressive probit models by including factors selected from a large number of predictors using principal component analysis and applies them to predict recessions in the United States. In addition, the factor-selecting process follows the proposed 'stepwise' procedure, which is customized for binary response variables. In terms of the empirical results, we should not impose the number of predictors or factors selected and extracting factors from the most informative predictors, rather than from all predictors, is more effective for recession predictions. Furthermore, the finding that different predictors or factors are most useful for different forecasting horizons implies that it is preferable not to use the same variables for different forecasting horizons.

Contents

1 Introduction

Recession predicting has become increasingly important in the last few decades, especially after the Global Financial Crisis (GFC) in 2007-2009. Whether the economy will be in an expansion or recession in the next few months provides significant information to policymakers, entrepreneurs and households, who need to make current decisions based on expectations of the future economic state.

Not surprisingly, there is a large literature related to recession forecasting. Due to restrictions of information technology, the early research mainly focuses on predicting recessions with a few identified predictors (for example, Stock & Watson, 1993; Estrella & Mishkin, 1998). Financial variables such as the slope of the yield curve and stock prices are the most used indicators for the prediction. Although these variables do have acceptable predictive power, the limitations of using very few regressors can severely erode the predictive performance.

Therefore, the idea of forecasting with a large set of predictors was introduced by Stock and Watson (2002b), who made use of principal component analysis to extract factors from a large panel of predictors. Since the factors contain the most essential information in the large panel, the proposed factor model outperforms some benchmark models such as univariate autoregression models. Instead of predicting real-valued economic activities in Stock and Watson (2002b), we intend to apply the methodology to recession forecasting. Although similar methods have been used to predict recessions (see Chen et al., 2011), our models are distinct from previous literature by including autoregressive terms and non-linear components.

In the applied econometrics literature, the probit model is most widely used to predict recessions, especially for U.S. recessions. In the early stage, the linear setting within the probit model only involves a static representation of regressors. Then, it is gradually extended to the dynamic autoregressive specification in Kauppi and Saikkonen (2008), which not only incorporates lagged values of driving predictors and the binary response variable but also includes lagged values of the linear specification itself. Figure 1, which is the out-of-sample prediction from Kauppi and Saikkonen (2008), illustrates that the estimated probabilities

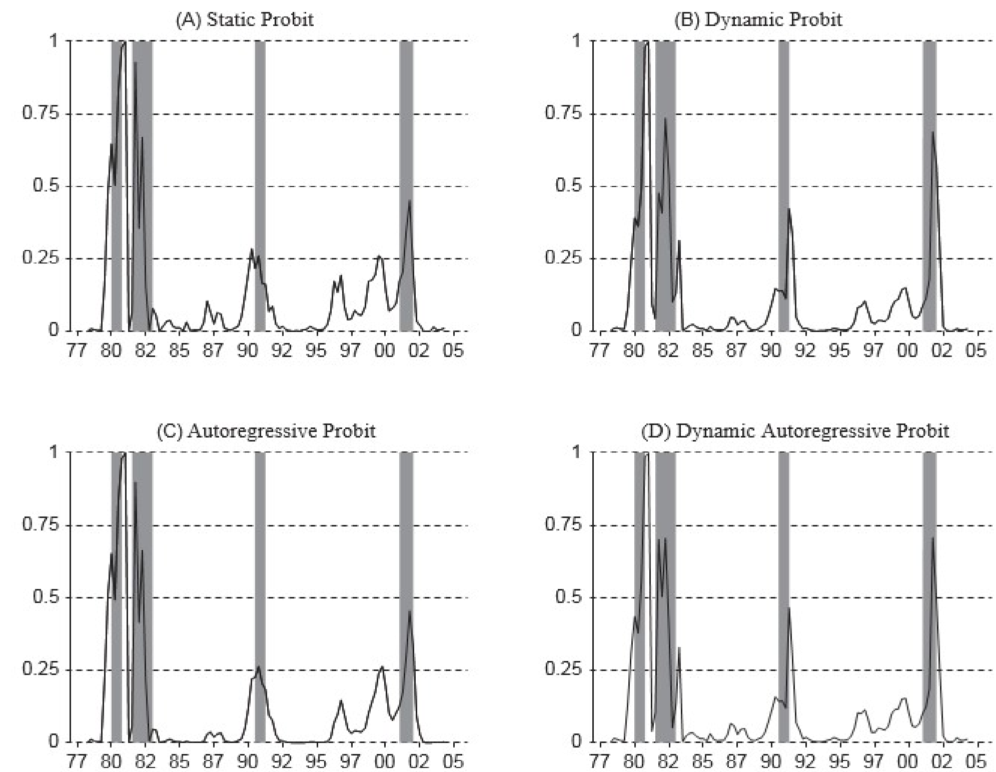

Figure 1. Probability of recession two quarters ahead, out-of-sample prediction. Adapted from "Predicting US recessions with dynamic binary response models", by H. Kauppi and P. Saikkonen, 2008, *The Review of Economics and Statistics, 90*(4), p. 785. Copyright 2008 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

based on the dynamic autoregressive pobit model generally matches better with the realized recessions than other three models. Nevertheless, the predictive power of that model is not strong enough for some recessions, which implies the possibility of model improvements. In particular, we apply our proposed models to forecast recessions over the period from late 1970s to early 1980s, which includes two apparent recessions.

Hence, we primarily extend Kauppi and Saikkonen's (2008) dynamic autoregressive probit model by replacing the interest rate spread with factors selected from a large number of predictors. We consider the case of selecting factors from all predictors as a simple benchmark. Alternatively, we extract factors from a targeted set of predictors selected by the Bayesian Information Criterion (BIC). It has been demonstrated in Bai and Ng (2008) that the forecasting performance for real-valued variables with the use of targeted predictors is superior. Therefore, we mainly focus on applying the methodology to a binary response variable and evaluating its helpfulness for recession prediction. Although the Least Absolute Shrinkage Selection Operator (LASSO) (Tibshirani, 1996) could be adapted as a selection criterion, the difficulty of finding a criterion with moderate computational complexity to determine the tuning parameter makes it less preferable. By changing the scenario from linear regression models to generalized linear models, we would need to adopt the penalized likelihood approach (Fan & Li, 2001) to select targeted predictors. Particularly, we investigate selecting predictors using BIC, which is widely used and bears little critique on computational expensiveness. In addition to the linear specification, we also consider models including squares of the predictors, and other nonlinear functions could be incorporated in the same way. This can be regarded as another contribution to the model construction.

In empirical analysis, We apply the proposed models to predict recessions in the United States and compare different models that use different predictors or factors by measuring their in-sample and out-of-sample performance. The results show that we should not impose a specific number of predictors when we only use individual predictors as regressors. In addition, extracting factors from a set of most informative predictors, rather than all predictors, improves the forecasting performance. For different forecasting horizons, the most useful variables are quite different. Therefore, we should allow different variables to be used for different forecasting horizons, rather than the same that Kauppi and Saikkonen (2008) suggested.

The remainder of this thesis is structured as follows. The relevant literature is reviewed in Section 2. The adopted framework for dynamic probit models and estimation procedures are introduced in Section 3. The simulation study is discussed in Section 4. Data and the empirical work are described and evaluated in Section 5 and Section 6, respectively. Section 7 concludes.

2 Related Literature

Over the last three decades, greater interest has been drawn to predicting recessions rather than only forecasting key quantities of economic activity. Two bodies of literature mainly contribute to this research field.

The first related literature is concentrated on identifying concrete recession predictors and constructing prediction models with a limited set of informative variables. Stock and Watson (1989, 1993) developed a recession prediction model in conjunction with the leading and coincident indicators. Their approach was to construct a VAR model and to derive the recession probabilities that it implies. One attractive feature of the model is that it empirically demonstrates the importance of financial variables such as interest rate spreads as leading indicators. This supports the view that financial variables provide significant information about whether an economy will be in a recession at a specific future date. More importantly, it incited a widespread interest in empirical research on the usefulness of financial variables for predicting the future state of an economy. For instance, Estrella and Mishkin (1998) focused on the out-of-sample performance of several financial variables as predictors of U.S. recessions. To avoid the relatively involved computation of recession probabilities implied by a high-dimensional vector autoregression (Stock & Watson, 1989, 1993), Estrella and Mishkin (1998) proposed to use the probit model, which directly predicts the binary dependent variable – whether the economy is in a recession or not.

One limitation of the linear relationship they defined within the probit model is only the values of selected variables at the predicted time are employed to forecast recessions.

This implies that no information about past economic status is used to form predictions. Several papers discussed and demonstrated the importance of including the lagged values of the binary dependent variable within the probit function. Dueker (1997) and Moneta (2005) argued that since the error terms are generally auto-correlated, one underlying assumption of the probit model that the random shocks are independent and identically distributed normal variables is violated. They showed that incorporating lag terms of the binary dependent variable not only enhances the validity of the random sampling assumption on the error terms, but also improves the forecasting power via including the information of past economic state.

More recently, Kauppi and Saikkonen (2008) extended previous dynamic probit models as follows in two innovative ways

$$\begin{aligned} \mathbb{E}_{t-1}(y_t) &= \mathbb{P}_{t-1}(y_t = 1) = \Phi(\pi_t) \\ &= \Phi\left(\sum_{j=1}^{p} \alpha_j \pi_{t-j} + \sum_{j=1}^{q} \delta_j y_{t-j} + x_{t-1}^{'}\beta + y_{t-d}x_{t-1}^{'}\gamma\right), \end{aligned} \tag{1}$$

where y_t and x_t represent the recession indicator and the interest rate spread respectively, π_t is determined by a considered specification, $\mathbb{E}_{t-1}(\cdot)$ and $\mathbb{P}_{t-1}(\cdot)$ denote as conditional expectation and conditional probability given the information set available at time $t-1$, and $\Phi(\cdot)$ is the cumulative distribution function of a standard normal distribution. Apart from including lags of the binary dependent variable (y_{t-j}) and lagged values of the interest rate spread (x_{t-1}), the authors also included the conditional probability of the dependent variable based on lagged values of π_t. This enriches the dynamic structure of π_t. Moreover, the interaction term $y_{t-d}x'_{t-1}$ represents the hypothesis that the effect of recession predictors would depend on a preceding state of the economy, effectively introducing a non-linear component. This model is the starting point for our work.

The second related literature, which has not been widely exploited, concerns using a large number of predictors to forecast recessions. Stock and Watson (1998) first introduced the idea of using a large set of predictors to forecast a macroeconomic time series variable rather than to predict an economic state of the business cycle. From the point of view of having a better in-sample results, it seems plausible to directly include a large number of predictor variables within a model. However, apart from the computational difficulty in including abundant explanatory variables, this is prone to resulting in a particularly poor out-of-sample performance due to the problem of overfitting. Hence, Stock and Watson (2002a, 2002b) developed a dynamic factor model. A handful of factors are extracted from all candidate predictors by principal component analysis, which approximately represent the most significant information containing in the large set of predictors. This innovation slacks the restriction that only a few series can be used in the previous forecasting process and the predictions based on a few factors perform well compared to several traditional benchmarks such as univariate autoregressive forecasts and vector autoregressive forecasts. Within the new proposed model, the authors only examined the linear predictive power of selected factors to an objective economic activity.

Consequently, one possible improvement in the model specification is to incorporate the predictors in a non-linear way. Bai and Ng (2008) suggested two feasible approaches to capture the non-linearity. Firstly, instead of having a linear relationship between the predictors and the factors, the method of quadratic principal components enriches the set of predictors by adding some or all cross-products of the predictors. Accordingly, extra non-linear characteristics of the predictors contribute to constructing the factors. These authors find that including all cross-products tends to lead to overparameterization issues and suggest that extracting factors from only the original predictors and their squares is sufficient. We follow their recommendation in our approach. Moreover, the authors improved the prediction procedure by considering the individual properties of a specific target series. This implies that the set of predictors used to extract the principal components is not necessarily the same for all economic series that we intend to forecast. Given a concrete series, they imposed thresholding rules such as LASSO and the elastic net to select the most informative predictors. With the use of these targeted predictors, forecasting improvements were realized over all predetermined forecast horizons from one month up to two years.

Although the common factor model mentioned above is usually applied to predict continuous real-valued macroeconomic variables (Stock & Watson, 2002a, 2002b; Bai & Ng, 2002, 2008), it is conceivable to expand its application to the forecasting framework of binary dependent variables such as recession indicators. Probit models are widely used to predict the probabilities of recessions (Dueker, 1997; Estrella & Mishkin, 1998; Chauvet & Potter, 2005; Moneta, 2005; Wright, 2006; Kauppi & Saikkonen, 2008), and most of the recent research related to recessions forecasting with common factors adopts the probit setup (for example, Chen et al., 2011; Christiansen et al., 2014). Chen et al. (2011), who first combined the probit model with common factors to predict U.S. recessions, formed their probit-dynamic factor model only employing the current values of common factors as the regressors. They demonstrated that their model outperforms several extant models, including the Estrella-Mishkin (1998) model and the Wright (2006) model, according to both in-sample and out-of-sample criteria. The superior performance is mainly attributed to the fact that more essential information has been embedded in the common factors. Moreover, problems associated with structural changes and data revisions in individual explanatory variables are mitigated with the use of common factors.

Apart from principal component regression, Bayesian shrinkage is regarded as another valid methodology to conduct predictions with a large panel of predictors. De Mol et al. (2008) proposed a scenario equivalent to a ridge regression to shrink the parameters of the regressors towards zero. Following their idea, Fornaro (2016) forecasted U.S. recessions with probit models estimated via Bayesian shrinkage. Compared to estimating latent factors by principal component analysis, Bayesian shrinkage improves the economic interpretation of the predictors. However, the computational expensiveness of using Bayesian methods, as well as their heavy reliance on somewhat arbitrary priors and the difficulty of incorporating nonlinearities in this setup, makes them less attractive for our purpose. Hence, we adopt principal component analysis combined with the BIC as the primary method to deal with high-dimensional data.

3 Models and Methodology

3.1 Models

Consider the binary-valued stochastic process $\{y_t\}_{t=1}^{T}$ which reflects the state of the economy as follows

$$y_t = \begin{cases} 1, & \text{if the economy is in a recession at time t} \\ 0, & \text{if the economy is not in a recession at time t} \end{cases} \tag{2}$$

Let $\mathcal{F}_t = \sigma\{(y_s, X_s), s \leq t\}$ be the information set available at time t, where X_t is an $N \times 1$ vector representing the data set of a large number of predictors at time t. Conditional on the information set $\mathcal{F}_{t-1}$, the recession indicator y_t has a Bernoulli distribution with probability p_t, that is, $y_t|\mathcal{F}_{t-1} \sim \mathcal{B}(p_t)$. In order to model the conditional probability p_t, we employ a probit model in which the conditional probability for a recession satisfies

$$\mathbb{E}_{t-1}(y_t) = \mathbb{P}_{t-1}(y_t = 1) = \Phi(\pi_t) = p_t \tag{3}$$

where $\mathbb{E}_{t-1}(\cdot)$, $\mathbb{P}_{t-1}(\cdot)$, $\Phi(\cdot)$ and π_t are defined as same as those in the equation (1). The setting of standard normal cumulative distribution function for $\Phi(\cdot)$ guarantees that the conditional probability p_t can only take values in the unit interval [0, 1]. Particularly, we consider the following form as the underlying specification of π_t

$$\pi_t = \omega + y_{t-1}\alpha + \pi_{t-1}\delta + f'_{t-1}\beta \tag{4}$$

where f_{t-1} is a general form representing a vector of common factors extracted from all predictors or targeted predictors selected by BIC. Each factor may be included at different lag orders, as advocated by Kauppi and Saikkonen (2008). We retain the notation f_{t-1} for simplicity but in practice, we always use lag lengths greater than or equal to the forecasting horizon. Here we do not include the interaction term such as $y_{t-1}x'_{t-1}\gamma$ mentioned in the equation (1) because we want to first keep the benchmark model simple. Besides, we introduce nonlinearity in a different way that is more standard in the factor model literature.

The idea of using targeted predictors is inspired by Bai and Ng (2008) who found improvements in predicting macroeconomic time series such as inflation by estimating common factors with fewer but most useful predictors. Given its notable predictive performance for continuous real-valued variables, we examine the performance if an analogous approach is applied to binary response variables. The elaboration of subset selection and parameters estimation will be presented in Section 3.2.

Given the benefits of including lagged values of the binary response variable and lagged values of π_t in Section 2, it is appropriate to incorporate them in the model specification of π_t. Additionally, Kauppi and Saikkonen (2008) illustrated that the first lag of both y_t and π_t have superior performance over other longer lags in both in-sample and out-of-sample forecasts. Hence, the model specification in equation (4) follows their findings. This model specification makes it possible to compare the predictive power of the new model and two existing models which are the dynamic autoregressive probit model proposed by Kauppi and Saikkonen (2008) and the probit-dynamic factor model introduced by Chen et al. (2011).

Compared to the former, the new model replaces the interest rate spread which is the only driving predictor by a few common factors; on the other hand, the new form extends the latter via including lagged values of y_t and π_t and provides refinements to the factor selection.

In addition to the linear specification of π_t, non-linear components could be considered to increase the forecasting accuracy. Accordingly, we allow f_{t-1} to depend on squares, which means that in this case f_{t-1} represents a general form of a vector of common factors selected from all predictors and their squared terms or a subset of all predictors and their squared terms. This means that we introduce the non-linearity by allowing the non-linear relationship between factors and predictors. Other non-linear functions may be considered, but good performance of models incorporating this form of nonlinearity is expected based on Bai and Ng (2008). Also, each factor may be extracted at different lag orders greater than or equal to the forecasting horizon.

3.2 Estimation of common factors and parameters

To extract common factors from a large set of predictors, we use principal component analysis advocated by Stock and Watson (2002a, 2002b) as the underlying methodology. Suppose we are given data on a large set of predictors $X_t = (X_{1t}, \ldots, X_{Nt})'$, where N is a large number and $t = 1, \ldots, T$. X_t is assumed to be $I(0)$ and we will further standardize all series with zero mean and unit variance when using real data. Suppose X_t satisfies the following equation

$$X_t = \Lambda F_t + e_t \tag{5}$$

where F_t is an $r \times 1$ vector of common factors, Λ is an $N \times r$ matrix of factor loadings, and $e_t = (e_{1t}, \ldots, e_{Nt})'$ is the $N \times 1$ idiosyncratic error. Given a static representation of the dynamic factor model such as (5), we can obtain consistent estimators of the latent factors F_t by principal component analysis[1] (Stock & Watson, 2002a). As we always work with $r \ll N$, a small number of factors effectively explain most of the variation in X_t. The most attractive advantage of applying principal component analysis is that it can reduce the high dimension of a large set of predictors without eroding the underlying information in the large panel. We employ these common factors in both the linear and the non-linear specification of π_t (see equation (4)), where f_{t-1} is selected from $\left(F'_{t-1}, \ldots, F'_{t-6}\right)'$ in both linear and non-linear structures. To determine the optimal set of factors, we first extract the first 10 common factors using principal component analysis and obtain a full set of factors with different lag orders between 1 and 6. In this case, we treat all 60 lagged factors as different individual variables; whether or not they should be included in the optimal set is determined by the following 'stepwise' procedure:

1. In the first step, consider 60 models, each of which includes only one of 60 lagged factors as f_{t-1} in the specification of π_t. Estimate all 60 models and pick the best one with the lowest value of BIC.

2. In the second step, consider 59 models, each of which is the model from the first step

[1] The principal component analysis problem can be simply described as:

$$\min_{F,\Lambda} ||X - F\Lambda^{'}||^2 \quad s.t. F^{'}F = I_{r\times r}$$

where X is a $T \times N$ matrix with rows $X_t^{'}$, F is a $T \times r$ matrix with rows $F_t^{'}$.
In the non-linear specification, X is replaced by (X, X^2), where the square is applied elementwise.

augmented with one of the remaining lagged factors. Estimate all 59 models and pick the best one with the lowest value of BIC.

3. In every next step, keep adding one of the remaining lagged factors at a time, stopping when the best model in a specific step has a larger BIC than the best model in the previous step.

The proposed 'stepwise' procedure is quite similar to the Least Angle Regression (LARS) algorithm, which is often used to yield the solution of LASSO under the Stagewise modification (Efron et al. 2004). However, given the binary response variable, we cannot directly apply the LARS algorithm; therefore, we introduce the above procedure to select factors that contribute to predicting recessions most. Besides, one may consider other more standard approaches such as panel information criteria; nevertheless, they aim at selecting factors fitting predictors best instead of the response variable. Moreover, this selection procedure maintains the idea of choosing factors that contain distinct information from each other by adding one more factor at each step. This probably keeps us from choosing factors having relatively strong collinearity.

As mentioned in Section 3.1, we are interested in selecting common factors from the most informative predictors rather than from all potential candidates. Bai and Ng (2008) proposed a selecting procedure by adopting soft thresholding methods such as LASSO to drop uninformative predictors. Although it is possible to apply a particular version of LASSO for generalized linear models (Fan & Li, 2001), the resulting optimization problem appears to be computationally intractable. More importantly, since the likelihood function is very flat in our case due to the limited amount of information contained in a binary response variable, so penalizing it in the way Fan and Li (2001) did would easily lead to the penalty being much more important than the information coming from the data itself, which is clearly undesirable. Hence, we select k^* predictors that bear most information following the selection procedure introduced above, employing the BIC which is not only easy to compute but also puts more penalty on overfitting.

For generalized linear models such as probit models, the estimation of parameters is usually carried out by maximizing likelihood functions. Suppose θ is a vector of parameters for π_t (for example, $\theta = \left(\omega, \alpha, \delta, \beta', \pi_0\right)'$ in (4)). Then, the log-likelihood function has the form

$$
\begin{aligned}
l(\theta) &= \sum_{t=1}^{T} l_t(\theta) \\
&= \sum_{t=1}^{T} \Big[y_t \log\big[\Phi(\pi_t(\theta))\big] + (1-y_t)\log\big[1-\Phi(\pi_t(\theta))\big]\Big],
\end{aligned}
\tag{6}
$$

where π_t is given by equation (4). After obtaining the maximum likelihood estimates, $\hat{\theta}$, we can calculate the BIC as

$$
BIC = -2 \cdot l(\hat{\theta}) + K \cdot ln(T) \tag{7}
$$

where K is the number of regressors within π_t, including the intercept. More specifically, suppose the optimal number of factors, r^*, is determined by the selection procedure, then K becomes $r^* + 4$ in this case.

In empirical analysis, we consider four underlying scenarios to determine factors: (A) choose the first five lagged predictors from all $\{x_{it}\}$ as factors directly; (B) choose k^* predictors from all $\{x_{it}\}$ selected by the 'stepwise' procedure as factors; (C) extract factors from all $\{x_{it}\}$ and determine r^* factors with their specific lag order chosen by BIC; (D) extract factors from the 30 best predictors in $\{x_{it}\}$ and determine r^* factors with their specific lag order chosen by BIC. For the first two scenarios, we directly select most relevant predictors as regressors without principal component analysis. The choices of five and 30 predictors, respectively, are based on Bai and Ng (2008). Other than selecting factors from all predictors themselves, we also choose factors from all predictors and their squares, $\{x_{it}, x_{it}^2\}$. Therefore, eight cases in total are considered for real data analysis; for each specification of π_t, we consider four scenarios respectively.

3.3 Forecasting procedures and evaluation measures

In this thesis, we predict U.S. recessions for $h = 3, 6, 12$ months ahead. We do not consider one-month-ahead forecasting since data updates for predictors usually have two months of delay. For instance, if it is currently June, we only have macro data until April while we are only interested in forecasts for July and later, so $h = 3$ is the minimal relevant horizon. Instead of relating the lags of regressors with the forecast horizon, we use the iterative approach elaborated in Kauppi and Saikkonen (2008). Given the law of iterated conditional expectations and the equation (3), the h months ahead forecast satisfies

$$\mathbb{E}_{t-h}(y_t) = \mathbb{E}_{t-h}\big(\mathbb{P}_{t-1}(y_t = 1)\big) = \mathbb{E}_{t-h}(\Phi(\pi_t)) \tag{8}$$

If $h = 3$, for example,

$$\mathbb{E}_{t-3}(y_t) = \sum_{y_{t-2}\in\{0,1\}} \sum_{y_{t-1}\in\{0,1\}} \left[\Phi(\pi_{t-2})\right]^{y_{t-2}} \times \left[1 - \Phi(\pi_{t-2})\right]^{1-y_{t-2}} \times \\ \left[\Phi(\pi_{t-1})\right]^{y_{t-1}} \times \left[1 - \Phi(\pi_{t-1})\right]^{1-y_{t-1}} \times \Phi(\pi_t) \tag{9}$$

Similar but more complicated procedures are applied to larger h.

We measure the in-sample performance of each model using monthly data from 1960:M1 to 2016:M3. For out-of-sample predictions, we use an estimation period containing 120 months for each out-of-sample forecast. In particular, if we are going to predict the economic state in January 1979, we use data from 1969:M1-h to 1979:M1-h for y_t and data from 1969:M1-2h-5 to 1979:M1-2h for macroeconomic variables to estimate models. This is because in the estimation period, we need to use h-month ahead macro data to fit data from 1969:M1-h to 1979:M1-h for y_t and we only allow six lag lengths starting from the number of forecasting horizon. By shifting the estimation period one month forwards at each time, we repeat the procedure over the out-of-sample forecasting period, which is from 1979:M1 to 1983:M12. We were planning to conduct forecasting over the period from 1972:M5 to 2016:M6; however, due to the time constraint, we only focus on the period from January 1979 to December 1983 in this thesis.[2] Although this period is about thirty years away from now, it covers two important U.S. recessions occurring over February 1980 to July 1980 and over August 1981 to November 1982 respectively. Therefore, it still remains meaningful to evaluate the forecasting performance for our proposed models under a relatively unstable

[2] The complete forecasting over 1972:M5 to 2016:M6 will continue after submitting this thesis.

state of the economy.

To evaluate the performance of the models both in-sample and out-of-sample, we adopt several evaluation metrics. The in-sample performance of the models is mainly assessed by the pseudo R^2 (Estrella, 1998), which is defined as

$$\text{pseudo } R^2 = 1 - \left(\frac{logL_u}{logL_c}\right)^{-(2/T)logL_c} \tag{10}$$

where L_u is the unconstrained maximum value of the likelihood function and L_c is the constrained maximum value of the likelihood function where we restrict the recession probability to be constant over time; that is, we replace equation (4) by $\pi_t = \pi_0$. Also, it is reasonable to evaluate models by other criteria such as BIC and compare results for different models. For the out-of-sample performance, we use the pseudo R^2 and the quadratic probability score (QPS)[3] to assess the model's forecasting accuracy. The former allows us to compare models with the dynamic autoregressive probit model defined in Kauppi and Saikkonen (2008); while the latter can be regarded as the counterpart of the mean squared forecast error in the real-valued models.

In this thesis, we do not test our results for significance. The main reason is that testing forecasting results is uncommon in this literature and it would be hard to come up with a useful test given how little information there is in a binary response variable.

[3] QPS is defined as

$$QPS = \frac{2}{T}\sum_{t=1}^{T}\left(\mathbb{E}_{t-h}(y_t) - y_t\right)^2$$

4 Simulation Study

4.1 Simulation procedures

Although it is impossible to know the data generating process (DGP) in the real world, we can mimic the process and assess the predictive ability of our proposed models by means of a Monte Carlo study. In this simulation study, we generate data of $\{y_t\}_{t=1}^T$ representing the economic state as follows

$$y_t = \begin{cases} 1, & \text{if } y_t^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where y_t^* is a latent variable and we specify it as

$$y_t^* = \pi_t + \varepsilon_t \tag{12}$$

where $\varepsilon_t \sim \text{WN}(0, 1)$. Since π_t follows both linear and non-linear specifications, y_t is generated according to the choice of π_t. Particularly, y_t follows a linear structure or a quadratic structure with the underlying π_t specified as equation (4).

After simulating the data of y_t, we extract factors and estimate parameters following the eight cases mentioned in section 3.2 for each DGP of y_t. For example, if y_t follows a linear structure, we first extract factors from all $\{x_{it}\}$ in four different ways and then repeat the same procedure by extracting factors from all $\{x_{it}, x_{it}^2\}$ instead. Using factors containing non-linearity to explain a linear response variable allows us to analyze what will happen if we use redundant information. Alternatively, if y_t actually follows a quadratic structure, using factors only selected from all predictors themselves can be regarded as one source of model misspecification. Since model misspecification commonly occurs in practice, it is worth mimicking the scenario and seeing to what extent it would affect forecasting performance for each of four distinct factor-selecting manners.

As mentioned in section 3.2, we allow lag orders of factors or selected predictors varying from 1 to 6. To be more realistic, the specific range of lag orders depends on one of three forecasting horizons. For instance, if the forecasting horizon is 3, the chosen factors or predictors should have lag orders between 3 and 8.

4.2 Simulation results

Although we are mainly interested in out-of-sample performance it is beneficial to first assess the in-sample performance of the proposed models. As mentioned in section 4.1, we consider eight cases for both linear and quadratic setting of y_t respectively.

Table 1 presents in-sample performance for eight scenarios of factors selection when y_t follows a linear structure. (A), (B), (C) and (D) are corresponding to four underlying ways that select predictors or factors (see section 3.2). For each case, the value of pseudo R^2 decreases as the forecasting horizon increases. This is expected because we have the most useful information related to the current period when conducting three-month ahead forecast. When we compare the four distinct methods of selecting predictors or factors, employing factors as regressors usually delivers a relatively poorer pseudo R^2 (see columns (C) and (D)). At first thought, these results seem a bit unreasonable since factors extracted by principal

Table 1: In-sample performance: y_t follows a linear structure

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h = 3$								
Pseudo R^2	0.7511	0.7448	0.6633	0.4945	0.7863	0.7863	0.6227	0.4529
BIC	97.5592	97.9990	110.1341	128.5138	91.2545	91.2545	117.1419	134.0166
$h = 6$								
Pseudo R^2	0.6846	0.6845	0.4131	0.3800	0.7431	0.7431	0.4056	0.3470
BIC	109.3210	109.1727	138.8301	140.4855	99.1384	99.1384	139.3266	143.2531
$h = 12$								
Pseudo R^2	0.6861	0.6861	0.4015	0.3782	0.7225	0.7225	0.3998	0.3566
BIC	109.1190	109.1190	140.1880	141.7912	102.9827	102.9827	139.3910	142.6919

component analysis are supposed to capture the most important information embedded in all available predictors. This means that once we restrict the number of regressors within a certain range,[4] the models using factors extracted from predictors or a targeted set of predictors should explain the variation in y_t better than the models including a few most relevant predictors. However, results shown in Table 1 do not verify this idea. This probably should be attributed to the reason that y_t is generated based on simulated predictors[5] rather than factors. Therefore, employing the most informative predictors would fit the simulated data better. Although these results do not match the ideal situation, it does stimulate our interest in seeing what will happen if we generate y_t with the use of factors. Besides, someone may be curious about why the values of pseudo R^2 and BIC are quite similar or exactly the same between cases (A) and (B). This is because when we allow models to freely choose predictors by the 'stepwise' procedure and set the maximum number of predictors chosen to

4 In this simulation study, we set the maximum number of predictors or factors to be selected as five to alleviate overfitting.

5 To reflect the properties of macroeconomic variables as much as possible, we generate 130 correlated variables following a first-order autoregressive (AR(1)) process.

be five, they usually end up with selecting the best five predictors, coinciding with case (A).

Other than comparing four cases within the group of factors selected either from $\{x_{it}\}$ or from $\{x_{it}, x_{it}^2\}$, it is important to compare results between groups for each case. By selecting factors from $\{x_{it}, x_{it}^2\}$ for y_t actually following a linear structure, we employ redundant information. This seems to be beneficial for choosing predictors directly as regressors because we now are able to include the squares of a more relevant predictor rather than include a less informative predictor. Hence, pseudo R^2 and BIC displayed in the second (A) and (B) in table 1 are all larger and smaller respectively than their counterparts in the first (A) and (B). Nevertheless, it turns out to be the other way around for factors selection. Presumably, redundant information in the squares of predictors makes it less effective to extract factors.

Moreover, table 2 displays in-sample performance for eight cases of factors selection when y_t follows a quadratic structure. Overall, the results are quite similar to those in table 1, no matter for the magnitudes of pseudo R^2 and BIC or for the comparative differences among four factor-selecting ways. However, the most essential difference is that for 3-month and

Table 2: In-sample performance: y_t follows a quadratic structure

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h=3$								
Pseudo R^2	0.7386	0.7386	0.6432	0.4343	0.7662	0.7654	0.6155	0.4416
BIC	99.7630	99.7630	112.2689	134.8682	94.8209	94.8147	118.2710	133.5724
$h=6$								
Pseudo R^2	0.6853	0.6853	0.3999	0.3625	0.7112	0.7089	0.3787	0.3492
BIC	108.8627	108.8627	140.3668	143.2926	104.6015	104.6204	141.6089	143.0389
$h=12$								
Pseudo R^2	0.6848	0.6848	0.4049	0.3379	0.7285	0.7285	0.3711	0.3667
BIC	109.1425	109.1425	140.8668	144.6043	101.7682	101.7682	142.3345	142.5751

12-month ahead forecasts factors extracted from the 30 best predictors in $\{x_{it}, x_{it}^2\}$ fit in-sample data better than those extracted from the 30 best predictors in $\{x_{it}\}$(compare pseudo R^2 and BIC in two columns of (D)). Although the absolute difference is not substantial, it seems to provide some evidence that model misspecification can somehow weaken in-sample fit.

Next we examine the out-of-sample forecasting performance of our proposed models. In this simulation study, we conduct 30 out-of-sample forecasts which can be treated as 30 months in reality.[6] As shown in table 3 and table 4, most values of pseudo R^2 are negative. This implies that regressors have a very poor out-of-sample fit since they are worse than a constant by itself. According to the results of QPS, we can make a similar conclusion since most values of QPS are not close to zero.[7] Even though the overall out-of-sample performance is disappointing, all positive values of pseudo R^2 occurring in the cases of employing factors is encouraging. Firstly, it demonstrates that factors extracted by principal component analysis do have stronger forecasting power than predictors when models almost include the same number of regressors. In other words, it is the most relevant information embedded in the selected factors that makes them superior to the best predictors. More importantly, it verifies the idea that a good in-sample fit does not guarantee a good out-of-sample performance. As shown in table 1 and table 2, no matter how we generate y_t or where we choose factors from, the selected predictors represented in (A) and (B) usually have a

[6] One possible concern is that the out-of-sample forecasting period is quite short. Due to the time constraint, we set the length of out-of-sample forecasting to be 30; but we will definitely try other longer period such as 100 later.

[7] The values of QPS range from zero to two with a value of zero corresponding to perfect accuracy. Specifically, it will yield QPS = 0.5 if we always predict that the probability of a recession is 0.5 in any month.

Table 3: Out-of-sample performance: y_t follows a linear structure

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h = 3$								
Pseudo R^2	-1.6138	-1.4186	0.4775	0.1064	-1.9087	-1.9087	0.3134	-1.1493
QPS	0.6526	0.6359	0.2735	0.4339	0.7868	0.7868	0.3292	0.7113
$h = 6$								
Pseudo R^2	-2.6300	-2.6264	-0.4865	-0.3084	-2.2064	-2.2064	-1.3965	-0.4476
QPS	0.8118	0.8114	0.6516	0.6346	0.6294	0.6294	0.7020	0.6774
$h = 12$								
Pseudo R^2	-1.0063	-1.0063	-0.7217	-0.4706	-1.4528	-1.4528	-1.0000	-0.6429
QPS	0.6125	0.6125	0.6490	0.6372	0.7307	0.7307	0.6638	0.7370

better in-sample performance. However, their out-of-sample fit is very poor, which probably results from overfitting. In contrast, the selected factors that have a relatively weak in-sample fit actually perform better in out-of-sample forecasts. This means that models with selected factors work better in out-of- sample forecasting despite the fact that our data don't even have a factor structure. Although this finding is meaningful, all four methods used to determine predictors or factors suffer from the issue of overfitting, which is more severe when we employ the selected predictors as regressors.

In addition to assessing in-sample and out-of-sample performance, we are interested in whether the proposed selection procedure choosing predictors or factors in correct lag orders. Since we generate y_t by using the fourth lag of the simulated predictors, selecting predictors or factors with the lag order of four is desired. Table 5, 6 and 7 display all chosen predictors or factors with their specific lag orders when predictors or factors are selected from $\{x_{it}\}$ with a linear y_t. Here, we only present the selection results for 3-month ahead forecasting as for

Table 4: Out-of-sample performance: y_t follows a quadratic structure

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h = 3$								
Pseudo R^2	-2.6692	-2.6692	-0.1304	0.2178	-3.0229	-3.0291	-0.2838	-0.0000
QPS	0.7407	0.7407	0.4252	0.3826	0.8273	0.8282	0.5375	0.4876
$h = 6$								
Pseudo R^2	-2.7631	-2.7631	-0.6067	-0.3724	-2.1951	-2.1576	-0.2185	-0.1833
QPS	0.8027	0.8027	0.6131	0.6452	0.8436	0.8319	0.5737	0.5978
$h = 12$								
Pseudo R^2	-2.0333	-2.0333	-1.3558	-0.6357	-1.9809	-1.9809	-0.3751	-0.1915
QPS	0.6964	0.6964	0.7045	0.6994	0.6796	0.6796	0.6333	0.5684

Table 5: Best 5 predictors selected from $\{x_{it}\}$: y_t follows a linear structure

h = 3											
	First 5 predictors						First 5 predictors				
Month	1st	2nd	3rd	4th	5th	Month	1st	2nd	3rd	4th	5th
1	30_{t-4}	22_{t-4}	27_{t-8}	108_{t-3}	99_{t-4}	16	22_{t-3}	6_{t-4}	71_{t-4}	17_{t-6}	36_{t-4}
2	28_{t-4}	41_{t-3}	44_{t-3}	75_{t-7}	109_{t-7}	17	8_{t-4}	18_{t-8}	38_{t-4}	7_{t-4}	69_{t-4}
3	7_{t-4}	16_{t-4}	115_{t-6}	41_{t-7}	45_{t-7}	18	1_{t-4}	5_{t-4}	24_{t-4}	73_{t-4}	58_{t-6}
4	9_{t-3}	42_{t-4}	5_{t-4}	14_{t-4}	4_{t-8}	19	34_{t-4}	67_{t-4}	71_{t-5}	11_{t-4}	32_{t-6}
5	29_{t-4}	25_{t-4}	39_{t-4}	18_{t-4}	57_{t-4}	20	29_{t-4}	10_{t-4}	89_{t-7}	93_{t-7}	33_{t-3}
6	8_{t-4}	56_{t-4}	122_{t-7}	14_{t-5}	73_{t-4}	21	15_{t-4}	88_{t-6}	10_{t-5}	5_{t-4}	9_{t-7}
7	105_{t-3}	66_{t-5}	61_{t-4}	41_{t-4}	23_{t-3}	22	14_{t-4}	91_{t-8}	10_{t-8}	77_{t-3}	118_{t-5}
8	25_{t-3}	54_{t-3}	10_{t-4}	10_{t-3}	6_{t-4}	23	8_{t-4}	29_{t-4}	97_{t-4}	8_{t-8}	104_{t-3}
9	5_{t-4}	28_{t-4}	57_{t-8}	54_{t-7}	15_{t-3}	24	55_{t-4}	3_{t-4}	113_{t-7}	5_{t-6}	109_{t-6}
10	69_{t-3}	19_{t-5}	120_{t-8}	130_{t-3}	117_{t-8}	25	23_{t-4}	35_{t-4}	118_{t-3}	12_{t-5}	87_{t-3}
11	73_{t-4}	60_{t-6}	8_{t-3}	85_{t-7}	5_{t-7}	26	16_{t-4}	31_{t-4}	77_{t-3}	115_{t-6}	87_{t-3}
12	8_{t-4}	40_{t-4}	26_{t-7}	6_{t-4}	22_{t-3}	27	16_{t-4}	54_{t-5}	38_{t-3}	8_{t-4}	10_{t-4}
13	45_{t-3}	43_{t-5}	80_{t-8}	116_{t-6}	130_{t-4}	28	10_{t-4}	57_{t-5}	9_{t-3}	25_{t-8}	50_{t-5}
14	6_{t-4}	44_{t-4}	127_{t-3}	52_{t-8}	66_{t-5}	29	40_{t-4}	32_{t-5}	128_{t-5}	112_{t-8}	87_{t-4}
15	105_{t-5}	26_{t-4}	69_{t-3}	10_{t-8}	21_{t-5}	30	46_{t-4}	5_{t-4}	18_{t-3}	79_{t-7}	72_{t-7}

other two longer forecasting horizons we only allow the models to select predictors or factors with the minimal lag order the same as the forecasting horizon. Hence, it is impossible to choose factors with fourth lag for 6-month and 12-month ahead forecasting. Besides, since the second factor-selecting method, choosing k^* predictors by the 'stepwise' procedure, often ends up with 5 predictors that are identical to those in the case of selecting best 5 predictors,

Table 6: Factors selected from $\{x_{it}\}$: y_t follows a linear structure

h = 3											
	Selected factors (at most five)						Selected factors (at most five)				
Month	1st	2nd	3rd	4th	5th	Month	1st	2nd	3rd	4th	5th
1	7_{t-4}	1_{t-4}	10_{t-7}	1_{t-7}	3_{t-5}	16	2_{t-4}	3_{t-4}	4_{t-8}	9_{t-5}	-
2	4_{t-4}	5_{t-4}	3_{t-4}	6_{t-4}	9_{t-4}	17	6_{t-4}	4_{t-4}	8_{t-4}	5_{t-5}	3_{t-4}
3	1_{t-4}	8_{t-4}	8_{t-8}	4_{t-7}	7_{t-4}	18	4_{t-4}	7_{t-4}	10_{t-4}	10_{t-3}	-
4	5_{t-4}	9_{t-4}	10_{t-4}	4_{t-3}	7_{t-3}	19	1_{t-4}	7_{t-4}	8_{t-4}	6_{t-4}	4_{t-4}
5	1_{t-4}	8_{t-3}	2_{t-5}	-	-	20	5_{t-4}	4_{t-4}	2_{t-4}	7_{t-8}	9_{t-4}
6	4_{t-4}	8_{t-4}	1_{t-4}	3_{t-4}	2_{t-6}	21	1_{t-4}	6_{t-4}	7_{t-8}	4_{t-4}	1_{t-7}
7	3_{t-4}	4_{t-4}	1_{t-4}	6_{t-3}	7_{t-3}	22	7_{t-4}	7_{t-8}	9_{t-3}	5_{t-4}	-
8	7_{t-4}	2_{t-4}	10_{t-4}	1_{t-4}	8_{t-4}	23	4_{t-4}	1_{t-4}	2_{t-4}	10_{t-4}	9_{t-4}
9	5_{t-4}	3_{t-4}	9_{t-4}	10_{t-4}	7_{t-4}	24	5_{t-4}	2_{t-4}	9_{t-5}	1_{t-4}	6_{t-4}
10	3_{t-4}	10_{t-4}	7_{t-4}	6_{t-4}	4_{t-3}	25	6_{t-4}	2_{t-4}	8_{t-4}	5_{t-4}	3_{t-3}
11	10_{t-8}	5_{t-4}	-	-	-	26	3_{t-4}	10_{t-6}	2_{t-4}	1_{t-8}	10_{t-5}
12	7_{t-4}	1_{t-4}	6_{t-4}	2_{t-3}	7_{t-6}	27	3_{t-4}	4_{t-4}	2_{t-8}	7_{t-4}	9_{t-4}
13	1_{t-5}	2_{t-5}	1_{t-6}	-	-	28	8_{t-4}	7_{t-4}	2_{t-4}	9_{t-4}	5_{t-5}
14	6_{t-4}	2_{t-4}	2_{t-3}	1_{t-4}	2_{t-6}	29	10_{t-4}	3_{t-5}	2_{t-4}	-	-
15	3_{t-4}	7_{t-4}	6_{t-4}	1_{t-4}	4_{t-5}	30	10_{t-4}	8_{t-4}	9_{t-3}	1_{t-5}	6_{t-5}

we do not display the result for this case. As shown in table 5 to 7, most of the first selected predictors or factors are in the lag order of four, which coincides with our expectation. In addition, the fourth lag is the most frequently selected lag order among the second to fifth selected predictors or factors. Based on these selection results, it is reasonable to conclude that the proposed 'stepwise' selection procedure does correctly choose predictors or factors that fit the response variable best. In particular, the cases shown in table 6 and 7 are corresponding to the first (C) and (D) in table 3, which deliver the values of out-of-sample

Table 7: Factors selected from a targeted set of $\{x_{it}\}$: y_t follows a linear structure

h = 3											
	Selected factors (at most five)						Selected factors (at most five)				
Month	1st	2nd	3rd	4th	5th	Month	1st	2nd	3rd	4th	5th
1	2_{t-4}	-	-	-	-	16	1_{t-4}	-	-	-	-
2	4_{t-3}	1_{t-4}	1_{t-5}	-	-	17	5_{t-3}	4_{t-3}	1_{t-6}	5_{t-7}	2_{t-8}
3	1_{t-4}	1_{t-7}	5_{t-4}	5_{t-6}	-	18	4_{t-3}	4_{t-4}	3_{t-4}	-	-
4	5_{t-4}	2_{t-4}	3_{t-4}	4_{t-4}	-	19	2_{t-4}	5_{t-6}	5_{t-3}	1_{t-3}	4_{t-5}
5	3_{t-3}	1_{t-8}	4_{t-4}	-	-	20	3_{t-4}	5_{t-4}	1_{t-7}	1_{t-4}	-
6	2_{t-4}	1_{t-4}	5_{t-4}	4_{t-6}	3_{t-6}	21	4_{t-4}	4_{t-6}	5_{t-3}	5_{t-4}	-
7	4_{t-4}	3_{t-4}	-	-	-	22	5_{t-8}	-	-	-	-
8	2_{t-4}	2_{t-7}	3_{t-4}	-	-	23	4_{t-4}	2_{t-4}	5_{t-6}	2_{t-8}	-
9	5_{t-4}	4_{t-4}	2_{t-4}	-	-	24	5_{t-4}	2_{t-4}	5_{t-3}	-	-
10	4_{t-3}	3_{t-3}	1_{t-5}	3_{t-8}	4_{t-4}	25	2_{t-4}	5_{t-4}	-	-	-
11	2_{t-3}	5_{t-7}	-	-	-	26	4_{t-4}	3_{t-4}	-	-	-
12	2_{t-7}	2_{t-4}	4_{t-4}	1_{t-4}	-	27	2_{t-4}	5_{t-4}	3_{t-3}	-	-
13	1_{t-4}	3_{t-4}	4_{t-3}	-	-	28	1_{t-4}	4_{t-3}	-	-	-
14	5_{t-4}	3_{t-6}	3_{t-3}	4_{t-8}	-	29	3_{t-3}	5_{t-7}	-	-	-
15	1_{t-4}	4_{t-5}	-	-	-	30	1_{t-8}	-	-	-	-

pseudo R^2 as 0.4775 and 0.1064 respectively. These two cases illustrate that their relatively better out-of-sample performance is likely attributed to a higher frequency of choosing the right lag order. Specifically, the proportion of selecting fourth lag is about 0.453 (68 out of 150) in table 5 while that is about 0.696 (96 out of 138) in table 6. A lower frequency of choosing right lag order may partially explain why the case corresponding to table 5 has the out-of-sample pseudo R^2 of -1.6138 (see table 3). Moreover, selecting a correct lag order seems to outweigh including more regressors, for determining out-of-sample performance. In terms of the results in table 7, two thirds of the out-of-sample forecasts select three or less than three factors and the proportion of choosing fourth lag order is 0.522 (47 out of 90).

Although the case of extracting factors from targeted predictors includes less regressors, it performs better in out-of-sample prediction than the case of choosing best 5 predictors.

In summary, although the cases with factors usually have a relatively weaker in-sample fit than those with individual predictors, they actually have a better out-of-sample forecasting performance. Additionally, our proposed 'stepwise' selection procedure works well in choosing correct predictors or factors and a higher frequency of selecting predictors or factors in right lag orders contributes to a better forecasting performance. Overall, the forecasting performance is not very good in most cases, and we suspect that it is because the BIC turns out to be an insufficient safeguard against overfitting. Hence, we hope to use alternative selection criteria such as cross-validation.

5 Data

We use data from two different sources to yield our empirical application. The first one is the U.S. business cycle expansion and contraction dates determined by the National Bureau of Economic Research (NBER)[8]. Following the definition of recessions described by Estrella and Trubin (2006), we use the data to determine U.S. recessions, which is indicated by y_t in our models. Figure 2 shows the plot of y_t from January 1960 to March 2016. Specifically,

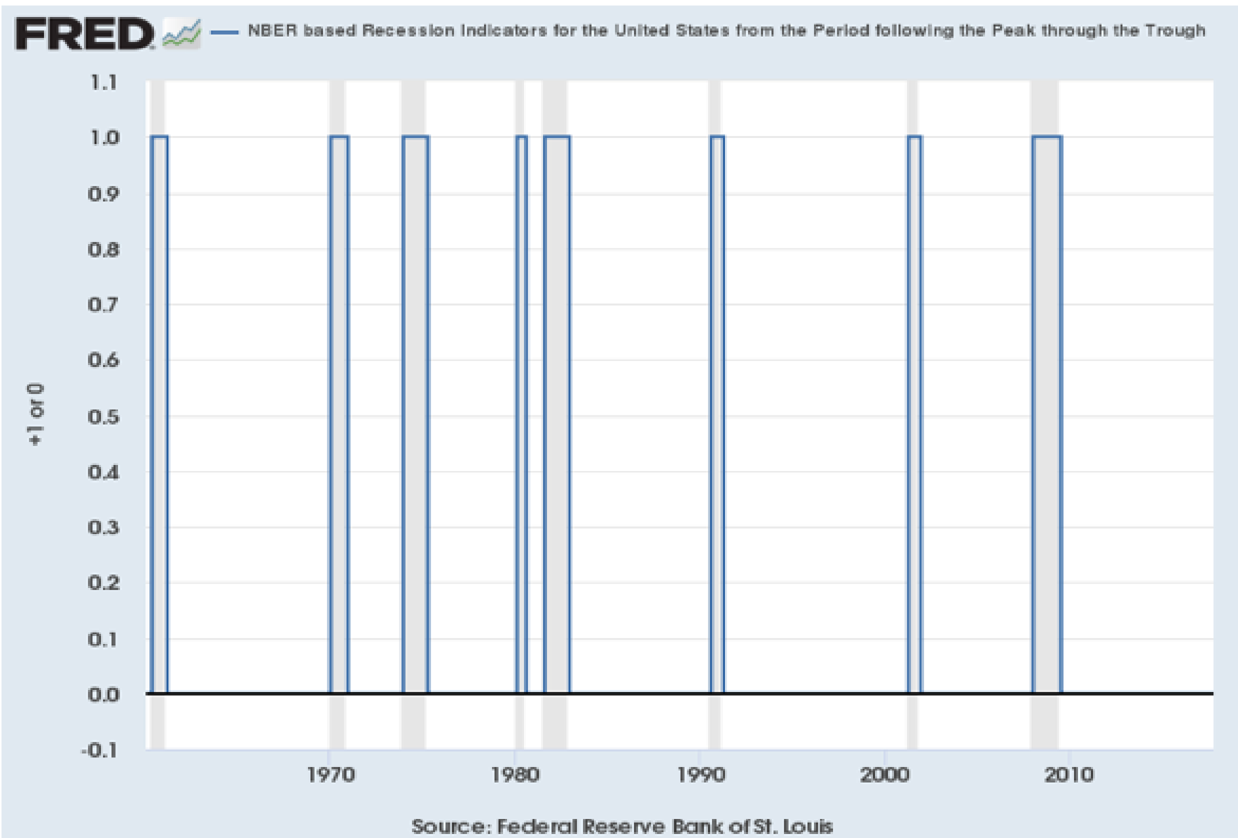

Figure 2: The plot of y_t from January 1960 to March 2016

the first month following a peak month is determined as the first recession month and the last month of a trough is determined as the last recession month. Another one is a

8 The business cycle dates are available at http://www.nber.org/cycles/cyclesmain.html.

large, monthly frequency macroeconomic dataset for predictive series, which is taken from McCracken's FRED-MD dataset.[9] The employed dataset spans the period from 1960:M1 to 2016:M3 and includes 123 variables[10] extracted from the FRED database. These variables, which form our x_{it}, cover broad categories of macroeconomic and financial series such as real activity indicators, interest rate indices and price indices.

[9] https://research.stlouisfed.org/econ/mccracken/fred-databases/

[10] Five variables were removed to obtain a balanced panel.

6 Empirical Analysis

As mentioned in Section 3.3, the in-sample estimation and performance are measured over the entire period from January 1960 to March 2016 while the out-of-sample predictions cover the period from January 1979 to December 1983. Since we do not need to generate data as we did for the simulation study, we only consider eight cases, four of which select predictors or factors from $\{x_{it}\}$ and four from $\{x_{it}, x_{it}^2\}$, as in the simulation. In addition, we transform McCracken's FRED-MD data to stationarity and remove outliers from the transformed data following the instruction on McCracken's website.[11]

6.1 Results

In empirical analysis, we still assess the in-sample fit first for each case. Table 8 presents in-sample performance and estimates for the constant, the coefficient of y_{t-1} and the coefficient of π_{t-1}, which are denoted as ω, α and δ in the model specification of π_t. Here we do not display coefficient estimates for selected regressors since the chosen regressors are quite different across eight cases, which implies that the comparative analysis among them seems useless. Besides, the standard errors are shown in parentheses, which are calculated following the idea of Estrella and Mishkin (1998). In particular, these standard errors are computed by applying the Newey-West (1987) technique with the consideration of the overlapping data issue.

Overall, $\hat{\omega}$ are negative for all three forecasting horizons across all eight cases, ranging

[11] https://research.stlouisfed.org/econ/mccracken/fred-databases/

Table 8: In-sample estimation and performance

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h = 3$								
$\hat{\omega}$	-5.7632	-5.7632	-3.5612	-2.6345	-5.0410	-5.0410	-2.3434	-2.4767
	(1.1943)	(1.1943)	(0.5651)	(0.2161)	(0.8769)	(0.8769)	(0.1630)	(0.1791)
$\hat{\alpha}$	7.5116	7.5116	4.9752	4.2810	7.0319	7.0319	3.4662	3.5670
	(1.4799)	(1.4799)	(0.7377)	(0.3765)	(1.1074)	(1.1074)	(0.2101)	(0.2141)
$\hat{\delta}$	-0.3670	-0.3670	-0.1421	-0.1060	-0.3540	-0.3540	0.0657	0.0885
	(0.0737)	(0.0737)	(0.0994)	(0.0633)	(0.1386)	(0.1386)	(0.0150)	(0.0058)
Pseudo R^2	0.8043	0.8043	0.7714	0.6942	0.8066	0.8066	0.7244	0.7505
BIC	117.7684	117.7684	125.6249	142.0091	116.7927	116.7927	117.1419	135.0532
$h = 6$								
$\hat{\omega}$	-3.6902	-3.6902	-2.6553	-2.6925	-3.4244	-3.4244	-2.9073	-2.5635
	(0.4608)	(0.4608)	(0.2201)	(0.2742)	(0.4123)	(0.4123)	(0.2510)	(0.1500)
$\hat{\alpha}$	5.6896	5.6896	4.0319	3.6979	5.4813	5.4813	4.3397	3.7539
	(0.7138)	(0.7138)	(0.3395)	(0.2943)	(0.6112)	(0.6112)	(0.3499)	(0.2494)
$\hat{\delta}$	-0.0647	-0.0647	0.0721	-0.1169	0.0061	0.0061	-0.2450	-0.0831
	(0.0761)	(0.0761)	(0.0203)	(0.0898)	(0.1279)	(0.1279)	(0.0579)	(0.0281)
Pseudo R^2	0.7815	0.7815	0.7347	0.6950	0.7915	0.7915	0.6756	0.6871
BIC	123.1801	123.1801	137.0587	142.2028	118.9892	118.9892	139.3266	139.4156
$h = 12$								
$\hat{\omega}$	-6.1744	-6.1744	-3.1371	-2.7921	-4.0264	-4.0264	-3.1615	-2.9190
	(2.0869)	(2.0869)	(0.4611)	(0.2809)	(0.4645)	(0.4645)	(0.4813)	(0.4671)
$\hat{\alpha}$	8.1399	8.1399	5.0841	4.1558	5.5357	5.5357	4.9019	4.5735
	(2.5658)	(2.5658)	(0.7337)	(0.4037)	(0.6705)	(0.6705)	(0.7729)	(0.5475)
$\hat{\delta}$	-0.5616	-0.5616	-0.2232	-0.1317	-0.2281	-0.2281	-0.2218	-0.2482
	(0.0538)	(0.0538)	(0.1250)	(0.0750)	(0.0945)	(0.0945)	(0.1570)	(0.0812)
Pseudo R^2	0.7746	0.7746	0.6884	0.6839	0.7495	0.7495	0.6928	0.6671
BIC	121.7836	121.7836	139.5608	141.5866	132.1922	132.1922	139.3910	142.7968

from -6.1744 to -2.3434. In contrast, $\hat{\alpha}$ are all significantly positive. This is expected because the current state of an economy should have a strong positive correlation with the economic state in the last time period. Compared to the coefficient estimate of y_{t-1} in Kauppi and Saikkonen's (2008) dynamic autoregressive probit model, our $\hat{\alpha}$ provides the same economic explanation by having a positive sign but on average has a larger value in magnitude. As for $\hat{\delta}$, although all of them reasonably have the absolute value within the unit interval (0, 1), most of them are undesirably negative. This also happened in Kauppi and Saikkonen's dynamic autoregressive probit model, but the most meaningful finding in our results is that there exist three statistically significant positive estimates for $\hat{\delta}$ occurring in the cases of extracting factors as regressors[12] (see the second (C) and (D) for $h = 3$ and the first (C) for $h = 6$ in table 8). This finding probably provides evidence that factors selected from predictors are more likely to capture the ambiguous variation in y_t so that the trend in π_{t-1} could be maintained in π_t.

For in-sample performance, it is quite similar to that discussed in the simulation study. Although the values of pseudo R^2 in the cases of selecting factors as regressors (see columns (C) and (D) in table 8) are still smaller than those in the cases of selecting predictors as regressors directly (see columns (A) and (B) in table 8), the absolute difference shrinks distinctly. This can be explained by the fact that the true generating process of y_t is really complicated in the real world. Hence, the relative better in-sample fit of employing predictors would not be as apparent as that found in the simulation study.

Next we would like to discuss the out-of-sample performance which is our main interest in. As shown in table 9, all cases have a pretty poor out-of-sample performance with negative pseudo R^2 and substantially larger value in magnitude than their counterparts in the simulation study. Since case (B) is usually better than case (A), we should not impose restriction

[12] In table 8, we can find another two positive values of $\hat{\delta}$ in the second (A) and (B) when $h = 6$. However, they are not statistically different from zero.

Table 9: Out-of-sample performance

	Factors selected from $\{x_{it}\}$				Factors selected from $\{x_{it}, x_{it}^2\}$			
	(A)	(B)	(C)	(D)	(A)	(B)	(C)	(D)
$h = 3$								
Pseudo R^2	-10.2809	-10.3132	-0.5051	-3.8669	-7.8083	-6.5805	-9.8860	-6.3026
QPS	0.8415	0.8460	0.2280	0.4701	0.7165	0.6415	0.8323	0.6971
$h = 6$								
Pseudo R^2	-9.9420	-7.4633	-11.6836	-9.7750	-9.4225	-7.4437	-9.3404	-9.3090
QPS	0.8165	0.7023	0.8998	0.9580	0.8082	0.6999	0.8031	0.8804
$h = 12$								
Pseudo R^2	-18.1247	-15.5654	-18.7240	-12.7126	-20.6074	-21.4355	-16.2446	-8.8546
QPS	1.1664	1.0581	1.1664	1.1062	1.2498	1.2747	1.0704	0.8896

on how many predictors are needed. Besides, the finding that case (D) is usually better than case (C) illustrates that the targeting procedure does help to improve the forecasting performance. Moreover, there is no clear evidence to decide between factors and regular predictors and thus no one can tell which of the four factor-selecting ways outperforming the other three in any forecasting horizon. It turns out that the issue of overfitting is aggravated when we apply to the real data. This probably redemonstrates the suspicion that BIC does not work well as expected to avoid overfitting, when there is very limited information in a binary response variable. As a result, we are going to try another selection criterion such as cross-validation later, which puts more penalty on overfitting.

As discussed in the simulation study, our proposed 'stepwise' selection procedure seems to select reasonable predictors or factors. Hence, we now evaluate whether it is still the case with the real data. Firstly, table 10 and 11 display the 10 most frequently selected predictors when we choose best five predictors from $\{x_{it}\}$ and from $\{x_{it}, x_{it}^2\}$, respectively.

All variables shown in these two tables are in the name of their mnemonics in FRED[13]

Table 10: 10 most frequently selected predictors: selecting best 5 predictors from $\{x_{it}\}$

	h = 3		h= 6		h= 12	
	Variable	Frequency	Variable	Frequency	Variable	Frequency
1	BAAFFM	29	COMPAPFFx	25	AWHMAN	16
2	HWIURATIO	19	AWHMAN	18	DMANEMP	16
3	AWHMAN	18	HOUST	13	IPNMAT	14
4	CES3000000008	8	UEMP15OV	13	PAYEMS	9
5	T10YFFM	7	HOUSTS	12	USTPU	9
6	HOUSTNE	6	USTPU	9	M1SL	7
7	AWOTMAN	6	NONREVSL	8	M2REAL	6
8	ISRATIOx	5	IPNCONGD	7	SRVPRD	5
9	INVEST	5	IPNMAT	5	RETAILx	5
10	USGOVT	5	CES2000000008	5	IPDMAT	5

Table 11: 10 most frequently selected predictors: selecting best 5 predictors from $\{x_{it}, x_{it}^2\}$

	h = 3		h= 6		h= 12	
	Variable	Frequency	Variable	Frequency	Variable	Frequency
1	AWHMAN^2	10	NDMANEMP^2	15	DMANEMP	15
2	UNRATE^2	9	AWHMAN^2	13	IPNMAT	14
3	NDMANEMP^2	9	UEMP15OV	8	AWHMAN^2	13
4	AWHMAN	8	CONSPI	8	PAYEMS	7
5	DMANEMP^2	7	HOUSTS^2	8	M2REAL	7
6	USGOOD^2	7	AWHMAN	7	USTPU	6
7	T10YFFM	6	COMPAPFFx	7	IPDMAT	6
8	WPSFD49207	5	USTPU	6	IPMANSICS	5
9	IPFPNSS^2	5	PERMITNE	4	DDURRG3M086SBEA^2	5
10	AAA	4	HOUSTW	4	M1SL	5

(also for table 12 and 13). For example, BAAFFM and HWIURATIO are the most two frequently chosen predictors when $h = 3$ in table 10. The former is an interest rate spread standing for 'Moody's Baa Corporate Bond Minus Federal Funds Rate' while the latter is a measure of labour market described as 'Ratio of Help Wanted/No. Unemployed'. Selecting them as the most related macroeconomic variables is appropriate since financial variables

[13] The specific description for each variable can be found in the document called "FRED-MD Updated Appendix" at https://research.stlouisfed.org/econ/mccracken/fred-databases/.

such as interest rate spreads or unemployment rates are commonly used as early indicators of changes in the state of an economy (Diebold, Rudebusch, & Aruoba, 2006; Stock & Watson, 1989, 1993). In addition, some selected predictors reflect the typical labour market structure responding to the estimation periods for out-of-sample predictions. For instance, frequently selected variables such as AWHMAN, DMANEMP, NDMANEMP and IPNMAT are all relevant to manufacturing industry.[14] This is because the estimation periods usually cover from late 1960s to early 1980s, and manufacturing industry dominated the labour market in the United States during that time (Lee & Mather, 2008). Although there was a labour shift away from manufacturing to services sector in 1970s and 1980s, manufacturing industry still accounted for a large share in the labour market.

Secondly, the cases of selecting optimal predictors from $\{x_{it}\}$ and from $\{x_{it}, x_{it}^2\}$ yield the similar results. As shown in table 12 and 13, variables such as BAAFFM, AWHMAN and NDMANEMP as well as their squares are still the most frequently selected variables. The main difference between these two cases and previous two cases is that the model often ends

Table 12: 6 most frequently selected predictors: selecting k^* predictors from $\{x_{it}\}$

	h = 3		h= 6		h= 12	
	Variable	Frequency	Variable	Frequency	Variable	Frequency
1	BAAFFM	29	COMPAPFFx	24	DMANEMP	15
2	AWHMAN	15	AWHMAN	18	AWHMAN	13
3	HWIURATIO	15	UEMP15OV	13	IPNMAT	12
4	T10YFFM	6	HOUSTS	10	PAYEMS	9
5	HOUSTNE	3	HOUST	9	CUSR0000SAD	5
6	BAA	3	USTPU	8	M2REAL	3

[14] AWHMAN: Avg Weekly hours - Manufacturing; DMANEMP: All Employees - Durable goods; NDMANEMP: All Employees - Nondurable goods; IPNMAT: Nondurable Materials.

Table 13: 6 most frequently selected predictors: selecting k^* predictors from $\{x_{it}, x_{it}^2\}$

	h = 3		h= 6		h= 12	
	Variable	Frequency	Variable	Frequency	Variable	Frequency
1	AWHMAN^2	9	NDMANEMP^2	15	DMANEMP	15
2	NDMANEMP^2	9	AWHMAN^2	12	AWHMAN^2	13
3	UNRATE^2	8	HOUSTS^2	8	IPNMAT	12
4	DMANEMP^2	7	COMPAPFFx	7	PAYEMS	7
5	AWHMAN	6	UEMP15OV	7	M2REAL	5
6	USGOOD^2	6	AWHMAN	6	IPDMAT	4

up choosing two predictors or sometimes including a third predictor. Nevertheless, including fewer predictors within π_t does not negatively affect the out-of-sample performance. Instead, it seems to slightly improve the out-of-sample fit if we compare the values of QPS in columns (A) and (B) in table 9, so our use of the BIC did alleviate the overfitting problem to some extent. However, both cases still exhibit a poor performance for out-of-sample predictions.

Thirdly, we assess the selection performance for the cases of extracting optimal factors from $\{x_{it}\}$ and from $\{x_{it}, x_{it}^2\}$. Table 14 and 15 show that the first principal component

Table 14: Frequency of each factor: extracting r^* factors from $\{x_{it}\}$

	h = 3		h= 6		h= 12	
	Factor	Frequency	Factor	Frequency	Factor	Frequency
1	4	60	6	47	1	45
2	8	21	8	32	9	26
3	9	9	2	29	5	18
4	6	6	10	21	10	17
5	10	2	4	6	4	13
6	2	1	3	5	6	11
7	3	1	9	5	7	8
8	5	1	1	5	8	5
9	7	1	5	5	3	2
10	1	0	7	3	2	1

Table 15: Frequency of each factor: extracting r^* factors from $\{x_{it}, x_{it}^2\}$

	h = 3		h= 6		h= 12	
	Factor	Frequency	Factor	Frequency	Factor	Frequency
1	3	37	7	37	1	48
2	6	33	1	24	5	21
3	10	15	4	24	7	17
4	8	13	8	21	8	16
5	1	13	6	19	10	9
6	4	12	3	18	9	5
7	5	12	9	7	2	2
8	9	10	10	7	3	1
9	7	8	2	3	4	0
10	2	0	5	2	6	0

is the most frequently selected factor when $h = 12$. This meets our expectation since the first principal component has the strongest correlation with all predictors, which basically measures general economic conditions. This further implies that in the longer run, the overall state of economy seems to be the most predictive. However, the first principal component is not quite frequently selected for other horizons, especially for $h = 3$ in table 14. Hence, other components appear to be more important at shorter horizons, which is consistent with the frequent selection of spreads and housing-related variables in table 12.

For the cases of extracting factors from a target set of $\{x_{it}\}$ or a target set of $\{x_{it}, x_{it}^2\}$, similar results are obtained in table 16 and 17. Since we now extract factors from the 30 most informative predictors rather than from 123 predictors, we reduce the number of principal components selected to be five. In these two cases, the first principal components are more frequently selected. This sounds reasonable as the target set of predictors is determined according to predictors' explanation power of the response variable y_t. Consequently, the

Table 16: Frequency of each factor: extracting factors from a target set of $\{x_{it}\}$

	h = 3		h= 6		h= 12	
	Factor	Frequency	Factor	Frequency	Factor	Frequency
1	2	36	1	32	1	49
2	3	36	2	31	4	26
3	5	31	4	25	5	24
4	1	22	3	24	2	16
5	4	21	5	19	3	13

Table 17: Frequency of each factor: extracting factors from a target set of $\{x_{it}, x_{it}^2\}$

	h = 3		h= 6		h= 12	
	Factor	Frequency	Factor	Frequency	Factor	Frequency
1	4	34	3	31	1	54
2	2	31	1	30	3	23
3	3	30	2	23	4	19
4	5	26	4	23	2	15
5	1	23	5	21	5	14

most related predictors have already been chosen in the target set so that the first principal component should be the most informative one to y_t. Obviously, there still exists the possibility that other factors are more frequently selected than the first one in practice (see $h = 3$ in both table 16 and table 17); and this can be also attributed to the reason that those factors put more weight on relatively more important predictors to y_t.

In summary, we have following five main conclusions about the empirical analysis. Firstly, based on the comparison between cases (A) and (B), we should not impose the restriction of always using five predictors as employing two or three predictors often works better. Secondly, extracting factors from a targeted set of predictors rather than from all predictors do improve the forecasting performance to some extent. Thirdly, there is no clear evidence to decide between individual predictors and factors, aiming to predict recessions better. Besides, since different variables are most useful for different horizons, we should not always use the same variables, which is contrary to what Kauppi and Saikkonen (2008) suggested. Last but not least, since our overall forecasting performance is not very good, we hope to improve this by using selection criteria that more closely mimic out-of-sample forecasting, such as cross-validation.

7 Conclusion

This thesis proposes new dynamic autoregressive probit models to predict U.S. recessions, with four distinct methods that determine which predictors or factors should be selected. Particularly, other than choosing individual predictors as regressors directly, we employ the principal component analysis to extract factors from all predictors or from the most informative predictors. Additionally, this thesis introduces a 'stepwise' selection procedure, which mimics the idea of the Least Angle Regression algorithm with the feasible version that can be applied to a binary response variable. The main contribution of this thesis is that we try to predict the state of an economy, which is a binary response variable, with a large panel of predictors using a more enriched model specification such as including autoregressive terms and nonlinearity.

By applying our proposed methodology to the simulated data, we illustrate that models with factors usually have a better forecasting performance than those with regular predictors, though factor models have a weaker in-sample fit. Moreover, the simulation results show that the proposed selection procedure possesses the ability to select reasonable predictors or factors for the binary response variable. According to our results in the empirical analysis, we should not restrict the number of predictors chosen to be five since including two or three predictors within the model often performs better. The results also demonstrate that constructing factors from a target set of predictors, rather than from all predictors, is more effective for recession predictions. Furthermore, the finding that different predictors or factors are most useful for different forecasting horizons implies that it is better not to use the same variables for different forecasting horizons. This is contrary to what was suggested by Kauppi and Saikkonen (2008).

Although this thesis provides some interesting findings, we acknowledge that our out-of-sample results still appear to suffer from overfitting problems. One possible way to alleviate this issue would be to use other model selection criteria that more closely mimic out-of-sample forecasting, such as cross-validation. Furthermore, one could consider other forms of nonlinearity, such as the interaction terms of the binary response variable and selected predictors or factors, in the style of Kauppi and Saikkonen (2008). Overall, we believe that our methodology promises to be a valuable addition to the toolkit of macroeconomic forecasters.

References

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317.

Chauvet, M., & Potter, S. (2005). Forecasting recessions using the yield curve. *Journal of Forecasting*, *24*(2), 77–103.

Chen, Z., Iqbal, A., & Lai, H. (2011). Forecasting the probability of US recessions: a Probit and dynamic factor modelling approach. *Canadian Journal of Economics/Revue canadienne d'économique*, *44*(2), 651–672.

Christiansen, C., Eriksen, J. N., & Møller, S. V. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, *49*, 459–468.

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*(2), 318–328.

Diebold, F. X., Rudebusch, G. D., & Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, *131*(1), 309–338.

Dueker, M. J. (1997). Strengthening the case for the yield curve as a predictor of U.S. recessions. *Federal Reserve Bank of St. Louis Review*, *79*, 41–51.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499.

Estrella, A. (1998). A new measure of fit for equations with dichotomous dependent variables. *Journal of Business & Economic Statistics*, *16*(2), 198–205.

Estrella, A., & Mishkin, F. S. (1998). Predicting U.S. recessions: financial variables as leading indicators. *Review of Economics and Statistics*, *80*(1), 45–61.

Estrella, A., & Trubin, M. (2006). The yield curve as a leading indicator: some practical issues. *Current Issues in Economics and Finance*, *12*(5).

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Fornaro, P. (2016). Forecasting US recessions with a large set of predictors. *Journal of Forecasting*, *35*, 477–492.

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

Lee, M., & Mather, M. (2008). *U.S. labor force trends* (Vol. 63) (No. 2). Population Reference Bureau.

Moneta, F. (2005). Does the yield spread predict recessions in the Euro area? *International Finance*, *8*(2), 263–301.

Newey, W. K., West, K. D., et al. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–08.

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. In *NBER Macroeconomics Annual 1989, Volume 4* (pp. 351–409). MIT press.

Stock, J. H., & Watson, M. W. (1993). A procedure for predicting recessions with leading indicators: econometric issues and recent experience. In *Business Cycles, Indicators and Forecasting* (pp. 95–156). University of Chicago Press.

Stock, J. H., & Watson, M. W. (1998). *Diffusion indexes* (Working Paper No. 6702). National Bureau of Economic Research.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179.

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Wright, J. H. (2006). The yield curve and predicting recessions. *Finance and Economics Discussion Series, Federal Reserve Board*.