Forecasting U.S. recessions with a large number of predictors

Shuting Chen

June 6, 2018

Abstract

This project introduces new dynamic autoregressive probit models by including factors (or predictors) selected from a large number of predictors with the use of principal component analysis (or adaptive boosting) and applying them to predict recessions in the United States. In addition, the factor/predictor-selecting process follows the proposed 'stepwise' procedure, which is customized for the binary response variable. In terms of the empirical results, determining factors or individual predictors with the 'stepwise' procedure performs in a reasonable way. 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.

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 (PCA) 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), I intend to apply the methodology to recession forecasting. Although similar methods have been used to predict recessions (see Chen et al., 2011), my 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. Although the estimated probabilities based on the dynamic autoregressive pobit model generally matches better, the predictive power of this model is not strong enough for some recessions, which implies the possibility of model improvements.

Hence, I 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. I consider the case of selecting factors by PCA from all predictors as a simple benchmark. Alternatively, I try to determine individual predictors by adaptive boosting, which would be helpful to increase the economic interpretations. In particular, I mainly focus on applying the methodology to a binary response variable and evaluating its helpfulness for recession prediction. By changing the scenario from linear regression models to generalized linear models, I would need to adopt the penalized likelihood approach (Fan & Li, 2001) to select targeted predictors. Specifically, I investigate selecting predictors using Bayesian Information Criterion (BIC), which is widely used and bears little critique on computational expensiveness. In addition to the linear specification, I 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, I 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 demonstrate that selecting factors (by PCA) or predictors (by adaptive boosting) with the proposed 'stepwise' selection precedure performs in a proper manner. 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. Data and the empirical work are described and evaluated in Section 4 and Section 5, respectively. Section 6 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 vector autoregression (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 forecast 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 this project.

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 PCA, 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 result in overparameterization issues and suggest that extracting factors from only the original predictors and their squares is sufficient. Therefore, this project follows their recommendation to include nonlinearlity by extracting factors from predictors and their squares.

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 existing 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. As a result, this project intends to predict recessions using common factors selected by PCA with a dynamic autoregressive probit model rather than the basic model used by Chen et al. (2011).

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, this project adopts principal component analysis combined with the BIC as the primary method to deal with high-dimensional data. Besides, this project would like to use adaptive boosting (Adaboost), which is an effective tool for classification problems (Ng, 2014), to identify several most important predictors for predicting recessions. Including specific predictors will not only enhance the economic interpretation of the model but also provide a chance to compare the forecasting performance of models using common factors with models using the same number of selected predictors.

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, this project employs 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 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, I 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 predictors determined by adaptive boosting. Each factor may be included at different lag orders, as advocated by Kauppi and Saikkonen (2008). I retain the notation f_{t-1} for simplicity but in practice, I always use lag lengths greater than or equal to the forecasting horizon.

Given the benefits of including lagged values of the binary response variable and lagged values of π_t (discussed in literature review), 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 or predictors; 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, I allow f_{t-1} to depend on squares, meaning that in this case f_{t-1} represents a general form of a vector of common factors selected from all predictors and their squares or predictors selected by adaptive boosting and their squares. This introduces the nonlinearity by allowing the nonlinear relationship between factors and predictors. Other nonlinear 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 (or predictors) and parameters

To extract common factors from a large set of predictors, I use principal component analysis advocated by Stock and Watson (2002a, 2002b) as the underlying methodology. 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. I employ these common factors in both linear and 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. Alternatively, one might be interested in a set of predictors which can predict recessions best. Hence, I adopt adaptive boosting (Ng, 2014) to determine the most informative predictors for recessions. This will not only enhance the economic interpretation of the model but also provide a chance to compare the forecasting performance of models using common factors with models using the same number of selected predictors.

To determine the optimal set of factors, I 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.[1] In this case, I 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.

[1] The same procedure is applied to the best 10 predictors determined by adaptive boosting.

2. In the second step, consider 59 models, each of which is the model from the first step 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, I cannot directly apply the LARS algorithm; therefore, I 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 the model from choosing factors having relatively strong collinearity.

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

$$\begin{aligned} l(\theta) &= \sum_{t=1}^{T} l_t(\theta) \\ &= \sum_{t=1}^{T} \Big[y_t \log\big[\Phi(\pi_t(\theta))\big] + (1 - y_t) \log\big[1 - \Phi(\pi_t(\theta))\big] \Big], \end{aligned} \tag{5}$$

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

$$BIC = -2 \cdot l(\hat{\theta}) + K \cdot ln(T) \tag{6}$$

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, I consider two underlying scenarios to determine factors or predictors: (A) extract factors from all $\{x_{it}\}$ and determine r^* factors with their specific lag order by the 'stepwise' procedure; (B) use adaptive boosting to determine a subset of best predictors related to recession forecasting and choose k^* predictors with their specific lag order using the 'stepwise' procedure. Other than selecting factors (or predictors) from all predictors themselves, we also choose factors (or predictors) from all predictors and their squares, $\{x_{it}, x_{it}^2\}$. Therefore, four cases in total are considered for real data analysis.

3.3 Forecasting procedures and evaluation measures

In this project, I predict U.S. recessions for $h = 3, 6, 12$ months ahead. I 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, I 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, I 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

$$\mathbb{E}_{t-h}(y_t) = \mathbb{E}_{t-h}\big(\mathbb{P}_{t-1}(y_t = 1)\big) = \mathbb{E}_{t-h}(\Phi(\pi_t)) \tag{7}$$

If $h = 3$, for example,

$$\begin{aligned}\mathbb{E}_{t-3}(y_t) = \sum_{y_{t-2}\in\{0,1\}} \sum_{y_{t-1}\in\{0,1\}} & \big[\Phi(\pi_{t-2})\big]^{y_{t-2}} \times \big[1-\Phi(\pi_{t-2})\big]^{1-y_{t-2}} \times \\ & \big[\Phi(\pi_{t-1})\big]^{y_{t-1}} \times \big[1-\Phi(\pi_{t-1})\big]^{1-y_{t-1}} \times \Phi(\pi_t)\end{aligned} \tag{8}$$

Similar but more complicated procedures are applied to larger h.

I measure the in-sample performance of each model using monthly data from 1960:M1 to 2017:M4 for macroeconomic variables and monthly data from 1960:M1 to 2018:M4 for y_t. For out-of-sample predictions, I 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, I 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, I need to use h-month ahead macro data to fit data from 1969:M1-h to 1979:M1-h for y_t and I only allow six lag lengths starting from the number of forecasting horizon. By shifting the estimation period one month forwards at each time, I repeat the procedure over the out-of-sample forecasting period, which is from 1972:M6 to 2018:M4.

To evaluate the performance of the models both in-sample and out-of-sample, I 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

$$\text{pseudo } R^2 = 1 - \Big(\frac{logL_u}{logL_c}\Big)^{-(2/T)logL_c} \tag{9}$$

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, I replace equation (3) 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, I use the pseudo R^2 and the quadratic probability score (QPS)[2] to assess the model's forecasting accuracy. The former allows me 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.

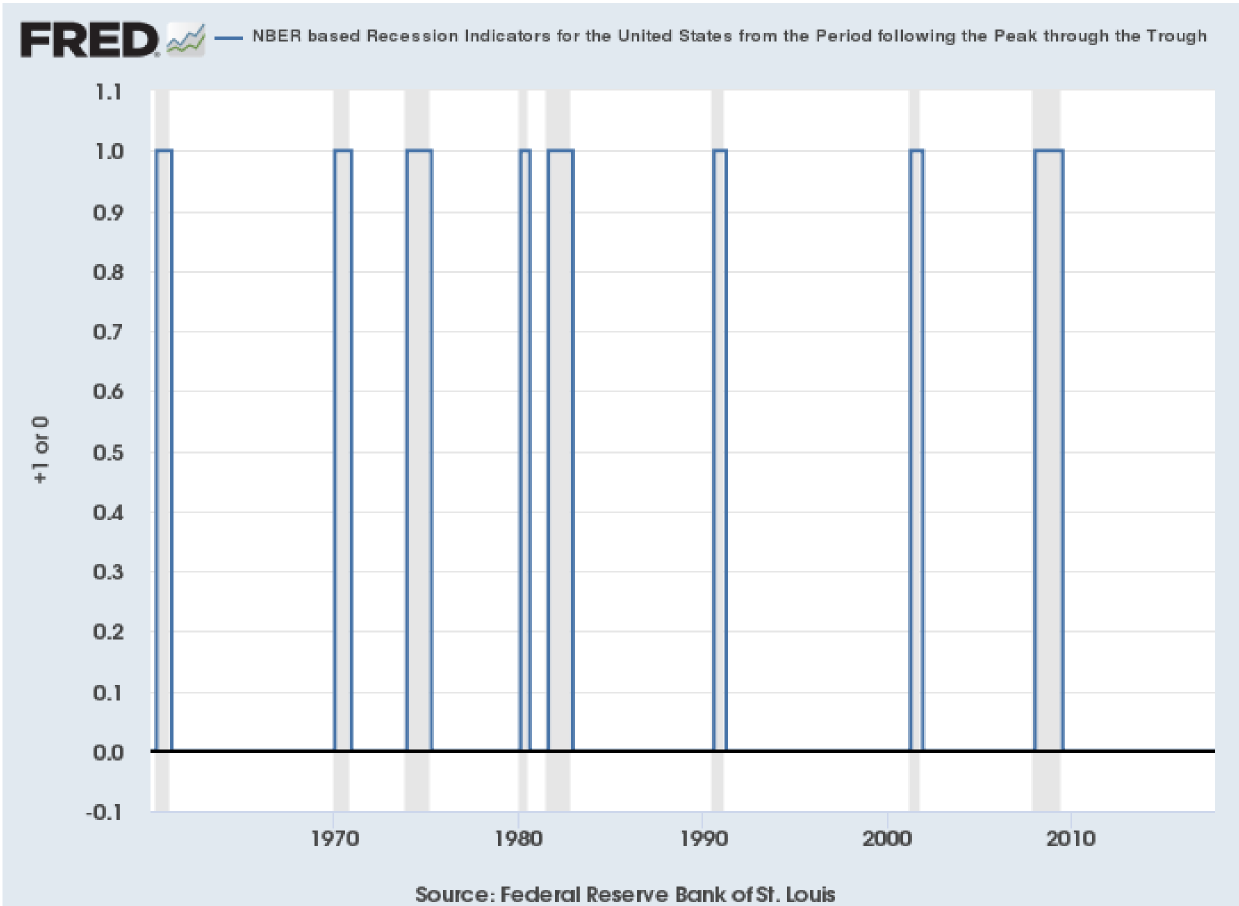
4 Data

I use data from two different sources to yield the empirical application. The first one is the U.S. business cycle expansion and contraction dates determined by the National Bureau of Economic Research (NBER)[3]. Following the definition of recessions described by Estrella and Trubin (2006), I use the data to determine U.S. recessions, which is indicated by y_t in my models. Figure 1 shows the plot of y_t from January 1960 to April 2018 with recession periods shaded in grey. Specifically, the first month following a peak month is determined as

[2] QPS is defined as

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

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



the first recession month and the last month of a trough is determined as the last recession month. Another one is a large, monthly frequency macroeconomic dataset for predictive series, which is taken from McCracken's FRED-MD dataset.[4] The employed dataset spans the period from 1960:M1 to 2017:M4 and includes 123 variables[5] extracted from the FRED database. These variables, which form x_{it}, cover broad categories of macroeconomic and financial series such as real activity indicators, interest rate indices and price indices.

4 https://research.stlouisfed.org/econ/mccracken/fred-databases/

5 Five variables were removed to obtain a balanced panel.

4.1 Summary statistics for key variables

As described in the online appendix of McCracken's FRED-MD dataset [6], the monthly macroeconomic series could be classified into eight groups, including output and income, labor market, interest and exchange rates, etc. Since the interest rate spread is often treated as the leading predictor of recessions, Table 1 displays the summary statistics for 10 macroeconomic series which are classified as interest and exchange rates.

Table 1: Summary Statistics for 10 Predictors
Classified as Interest and Exchange Rates

Fred	Description	Mean(%)	S. D.
FEDFUNDS	Effective Federal Funds Rate	5.1085	3.6843
CP3Mx	3-Month AA Financial Commercial Paper Rate	5.2132	3.4627
TB3MS	3-Month Treasury Bill	4.6322	3.1724
TB6MS	6-Month Treasury Bill	4.7704	3.1582
GS1	1-Year Treasury Rate	5.1491	3.3706
GS5	5-Year Treasury Rate	5.8484	3.0684
GS10	10-Year Treasury Rate	6.1815	2.8436
AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	2.0696	1.9733
BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	3.0836	2.0758
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	-0.4764	0.7146

As shown in Table 1, the average rate of federal funds or different maturity treasury bills/bonds are quite similar and almost around five percent. Besides, the standard deviations of these bonds remain at similar level. Since corporate bonds are riskier, they have higher expected return than federal funds. Additionally, the differences between corporate bonds and federal funds (AAAFFM and BAAFFM) experience less variation over time, with smaller standard deviations than pure federal funds/treasury bonds.

[6] https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf

5 Results

As mentioned in Section 3.3, the in-sample estimation and performance are measured over the entire period from January 1960 to April 2017 while the out-of-sample predictions cover the period from June 1972 to April 2018. Additionally, I transform McCracken's FRED-MD data to stationarity and remove outliers from the transformed data following the instruction on McCracken's website.[7]

Before evaluating the out-of-sample performance, I assess the in-sample fit first for each case. Table 2 displays all chosen factors (extracted by PCA) with their specific lag orders when they are selected from both $\{x_{it}\}$ and $\{x_{it}, x_{it}^2\}$. To avoid overfitting, I set the maximum

Table 2: Best selected common factors (at most five): in-sample

	Factors selected from $\{x_{it}\}$					Factors selected from $\{x_{it}, x_{it}^2\}$				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
$h = 3$										
Factors	4_{t-4}	2_{t-7}	7_{t-6}	6_{t-8}	-	6_{t-4}	3_{t-7}	-	-	-
$h = 6$										
Factors	2_{t-7}	6_{t-10}	-	-	-	3_{t-8}	-	-	-	-
$h = 12$										
Factors	1_{t-12}	9_{t-13}	-	-	-	8_{t-16}	1_{t-12}	-	-	-

number of factors included in the model to be five. As shown in Table 2, most models include one or two factors using the 'stepwise' selection procedure described above. For different forecasting horizons, the first selected factor is usually close to the forecasting horizon. Nevertheless, there is no strong pattern about which principal component among first 10 of them should be selected first.

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

Table 3 displays in-sample estimation and performance for common factors. $\hat{\omega}$, $\hat{\alpha}$ and $\hat{\delta}$ represent estimates for constant, y_{t-1} and π_{t-1} respectively. Overall, $\hat{\omega}$ are negative for all three forecasting horizons, ranging from -3.6224 to -2.3522. 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. As for $\hat{\delta}$, although all of them reasonably have the absolute value within the unit interval (0, 1), most

Table 3: In-sample estimation and performance: common factors

	Factors selected from $\{x_{it}\}$			Factors selected from $\{x_{it}, x_{it}^2\}$		
	$h = 3$	$h = 6$	$h = 12$	$h = 3$	$h = 6$	$h = 12$
$\hat{\omega}$	-3.6224	-3.1210	-3.1711	-2.3522	-2.9032	-3.1765
	(0.5909)	(0.4664)	(0.4553)	(0.1671)	(0.2471)	(0.4901)
$\hat{\alpha}$	5.0674	4.5593	5.1468	3.4761	4.3272	4.9118
	(0.7681)	(0.7174)	(0.7275)	(0.2107)	(0.3401)	(0.7857)
$\hat{\delta}$	-0.1600	-0.1915	-0.2260	0.0655	-0.2383	-0.2217
	(0.1072)	(0.1341)	(0.1201)	(0.0113)	(0.0585)	(0.1592)
Pseudo R^2	0.7681	0.7005	0.6862	0.7212	0.6722	0.6900
BIC	125.9426	138.3900	139.3141	134.4071	145.2145	137.5726

of them are undesirably negative except for one in the right panel when $h = 3$. Comparing coefficients in the right panel to those in the left panel of Table 3, we can conclude that including nonlinearity by selecting factors from $\{x_{it}, x_{it}^2\}$ makes estimates have smaller standard errors (shown in parentheses).

For in-sample performance, 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. Additionally, the value of BIC increases as the forecasting horizon goes up.

For predictors selected by adaptive boosting, Table 4 shows the overall in-sample estimation and performance. Compared to coefficients of $\hat{\omega}$, $\hat{\alpha}$, and $\hat{\delta}$ in Table 3, the estimates in Table 4 retain the same level and the same sign, though with a bit smaller magnitude.

Table 4: In-sample estimation and performance: predictors (Adaboosting)

	Predictors selected from $\{x_{it}\}$			Predictors selected from $\{x_{it}, x_{it}^2\}$		
	$h = 3$	$h = 6$	$h = 12$	$h = 3$	$h = 6$	$h = 12$
$\hat{\omega}$	-2.9987	-2.6258	-2.6596	-3.0019	-2.6291	-2.6614
	(0.4982)	(0.1784)	(0.3579)	(0.4985)	(0.1782)	(0.3581)
$\hat{\alpha}$	4.3502	3.6847	4.3814	4.3502	3.6847	4.3814
	(0.7114)	(0.2390)	(0.5652)	(0.7114)	(0.2390)	(0.5652)
$\hat{\delta}$	-0.2436	-0.1269	-0.1039	-0.2436	-0.1269	-0.1039
	(0.1884)	(0.0388)	(0.1273)	(0.1884)	(0.0388)	(0.1360)
Pseudo R^2	0.7065	0.6792	0.6686	0.7065	0.6792	0.6686
BIC	134.8733	141.8932	140.8571	134.8733	141.8932	140.8571

Although the overall trend of pseudo R^2 and BIC is very similar to that in Table 3 as forecasting horizon goes up, the values of pseudo R^2 (or BIC) in Table 4 are smaller (or larger) than their counterparts in Table 3. This implies that the models with common factors determined by PCA have a better forecasting performance than those with specific predictors chosen by adaptive boosting. One might be curious about why estimates are almost exactly the same between right panel and left panel in Table 4. This is because no matter how predictors are selected, whether from $\{x_{it}\}$ or from $\{x_{it}, x_{it}^2\}$, they always end up with the same predictors. More specifically, TB6SMFFM, which is 6 mon-FF(Treasury Bill minus Effective Federal Funds Rate) spread, is the most frequently selected predictor when $h = 3$ and 6.

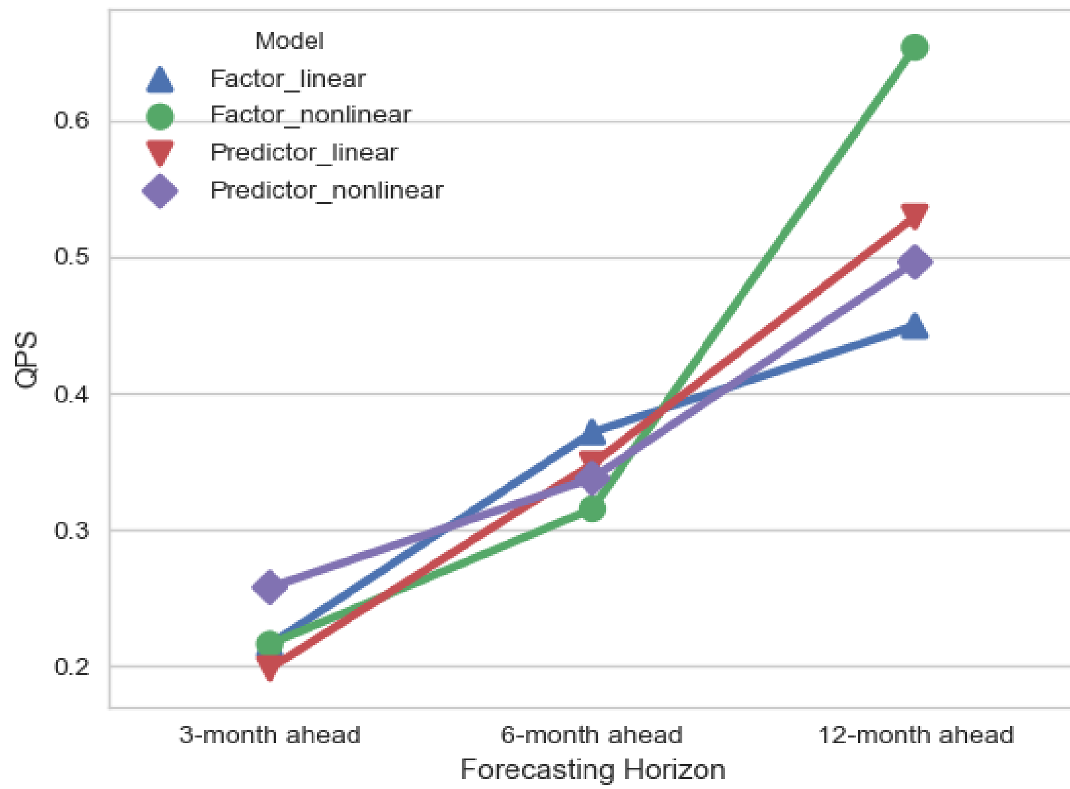
Next, I would like to discuss the out-of-sample performance which is my main interest in. As shown in Table 5, all cases have a poor out-of-sample performance with negative pseudo R^2 and larger value in magnitude than their counterparts in in-sample performance. The negative pseudo R^2 demonstrates the issue of over-fitting, however, it is hard to deliver meaningful explanations based on the negative pseudo R^2. Therefore, I adopt QPS (elaborated in section 3.3), which could be treated as the counterpart of the mean squared forecast error in the real-valued models to assess the out-of-sample forecasting accuracy. Based on the definition of QPS (see section 3.3), 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. As shown in Table 5 and Figure 2, the value of QPS goes up as the forecasting horizon increases for all four cases. This is reasonable and expected because the forecasting accuracy should be weakened as we have less relevant information about economy state with longer forecasting horizon.

Table 5: Out-of-sample performance

	Factors/Predictors selected from $\{x_{it}\}$		Factors/Predictors selected from $\{x_{it}, x_{it}^2\}$	
	(A) Factors	(B) Predictors	(C) Factors	(D) Predictors
$h = 3$				
Pseudo R^2	-0.9927	-0.8530	-1.0181	-1.2391
QPS	0.2156	0.1980	0.2159	0.2575
$h = 6$				
Pseudo R^2	-2.3136	-1.9074	-1.8425	-1.9771
QPS	0.3714	0.3480	0.3152	0.3376
$h = 12$				
Pseudo R^2	-2.7594	-3.2793	-4.0098	-3.0074
QPS	0.4495	0.5294	0.6539	0.4963

Moreover, there is no clear evidence to decide between factors and regular predictors and thus no one can tell which of the four factor/predictor-selecting ways outperforming the other three in any forecasting horizon. It turns out that the issue of over-fitting is salient for out-of-sample prediction. This probably demonstrates the suspicion that BIC does not work well as expected to avoid over-fitting, when there is very limited information in a binary

Figure 2: Out-Of-Sample Performance (QPS) for Four Models



response variable. As a result, it would be better to try another selection criterion such as cross-validation later, which puts more penalty on over-fitting.

Other than evaluating the out-of-sample performance, it is meaningful to find out whether adaptive boosting and the proposed 'stepwise' selection procedure select reasonable predictors. Table 6 displays 10 most frequently selected predictors when choosing the best five predictors from $\{x_{it}\}$. All variables shown in Table 6 are in the name of their mnemonics in FRED[8]. For example, TB6SMFFM and USGOOD are the two most frequently chosen predictors when $h = 3$. The former is an interest spread standing for '6-Month Treasury C Minus FEDFUNDS' while the latter is a measure of labor market described as 'All Em-

[8] 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/.

Table 6: 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	TB6SMFFM	87	TB3SMFFM	75	AWHMAN	48
2	USGOOD	57	TB6SMFFM	54	TB3SMFFM	48
3	IPMANSICS	54	MANEMP	39	IPNMAT	42
4	MANEMP	24	HOUST	39	IPMANSICS	27
5	TB3SMFFM	21	HWIURATIO	36	PERMIT	27
6	HWIURATIO	18	UEMP150V	27	M1SL	21
7	FEDFUNDS	18	IPMANSICS	24	M2REAL	18
8	CUMFNS	15	USGOOD	21	FEDFUNDS	15
9	PERMIT	15	IPNMAT	15	RETAILx	15
10	PERMITW	15	CES2000000008	15	MANEMP	15

ployees: Goods-Producing Industries'. Selecting them as the most related macroeconomic variables is appropriate since financial variables 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). Besides, after comparing selected predictors among three different forecasting horizons, it seems appropriate to conclude that individual predictors chosen for each forecasting horizon do not necessarily remain the same.

6 Conclusion

This project proposes new dynamic autoregressive probit models to predict U.S. recessions, with two distinct methods that determine which predictors or factors should be selected. Particularly, other than choosing individual predictors by adaptive boosting as regressors directly, I employ the principal component analysis to extract factors from all predictors. Additionally, this project 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 project is that I try to predict the state of an economy, which is a binary response variable, with a large panel of predictors using more enriched model specifications such as including autoregressive terms and nonlinearity.

According to the results in the empirical analysis, combining PCA or adaptive boosting with 'stepwise' selection procedure does determine factors or individual predictors in a proper way. 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 project provides some interesting findings, I acknowledge that my out-of-sample results still appear to suffer from over-fitting 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, I believe that my 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 us 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.

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

Ng, S. (2014). Boosting recessions. *Canadian Journal of Economics/Revue canadienne d'économique*, *47*(1), 1–34.

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* (No. w6702). 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.

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