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

Shuting Chen

May 9, 2018

1 Models and Methodology

1.1 Models

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

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

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

$$\mathbb{E}_{t-1}(y_t) = \mathbb{P}_{t-1}(y_t = 1) = \Phi(\pi_t) = p_t \tag{2}$$

where 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. 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

$$\pi_t = \omega + y_{t-1}\alpha + \pi_{t-1}\delta + f'_{t-1}\beta \tag{3}$$

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 (3) 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.

1.2 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 (3)), 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 (3)). Then, the log-likelihood function has the form

$$
\begin{aligned}
l(\theta) &= \sum_{t=1}^{T} l_t(\theta) \\
&= \sum_{t=1}^{T} \Big[y_t \log\big[\Phi(\pi_t(\theta))\big] + (1-y_t)\log\big[1-\Phi(\pi_t(\theta))\big]\Big],
\end{aligned} \tag{4}
$$

where π_t is given by equation (3). After obtaining the maximum likelihood estimates, $\hat{\theta}$, we can calculate the BIC as

$$
BIC = -2 \cdot l(\hat{\theta}) + K \cdot ln(T) \tag{5}
$$

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.

1.3 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 (2), the h months ahead forecast satisfies

$$\mathbb{E}_{t-h}(y_t) = \mathbb{E}_{t-h}\big(\mathbb{P}_{t-1}(y_t = 1)\big) = \mathbb{E}_{t-h}(\Phi(\pi_t)) \tag{6}$$

If $h = 3$, for example,

$$\begin{aligned}\mathbb{E}_{t-3}(y_t) = \sum_{y_{t-2}\in\{0,1\}} \sum_{y_{t-1}\in\{0,1\}} & \big[\Phi(\pi_{t-2})\big]^{y_{t-2}} \times \big[1 - \Phi(\pi_{t-2})\big]^{1-y_{t-2}} \times \\ & \big[\Phi(\pi_{t-1})\big]^{y_{t-1}} \times \big[1 - \Phi(\pi_{t-1})\big]^{1-y_{t-1}} \times \Phi(\pi_t)\end{aligned} \tag{7}$$

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

$$\text{pseudo } R^2 = 1 - \left(\frac{logL_u}{logL_c}\right)^{-(2/T)logL_c} \tag{8}$$

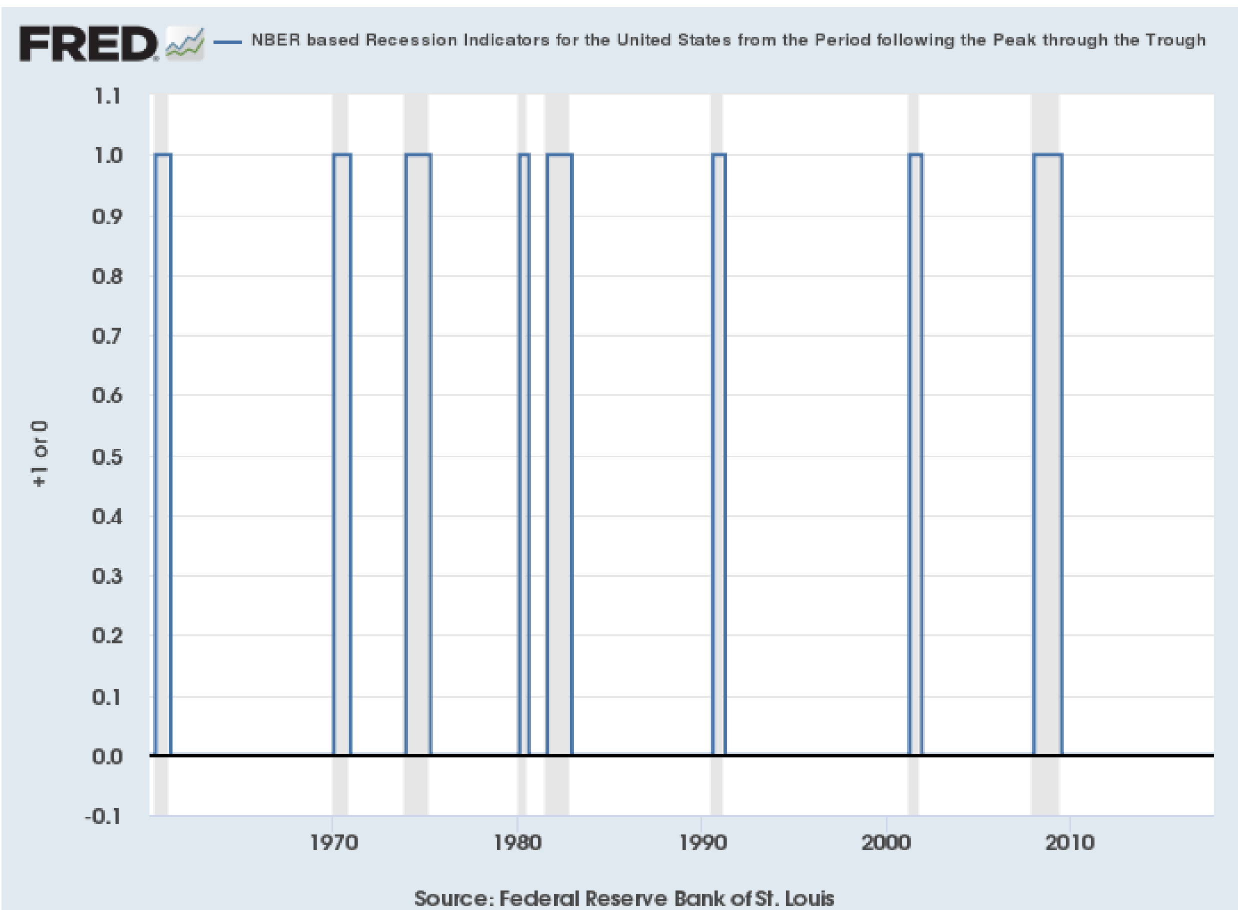
where L_u is the unconstrained maximum value of the likelihood function and L_c is the constrained maximum value of the likelihood function where we restrict the recession probability to be constant over time; that is, I replace equation (3) by $\pi_t = \pi_0$. Also, it is reasonable to evaluate models 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.

[2] QPS is defined as

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

2 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



the first recession month and the last month of a trough is determined as the last recession

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

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.

2.1 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

[4] https://research.stlouisfed.org/econ/mccracken/fred-databases/
[5] Five variables were removed to obtain a balanced panel.
[6] https://s3.amazonaws.com/files.fred.stlouisfed.org/fred-md/Appendix_Tables_Update.pdf

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.

3 Preliminary results

By now, I have obtained in-sample estimation and performance for factors extracted by principal component analysis and still working on the case for predictors determined by adaptive boosting. After finishing in-sample estimation, I will move on to the out-of-sample estimation and performance evaluation.

Table 2 displays all chosen factors with their specific lag orders when they are selected

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}	-	-	-

from both $\{x_{it}\}$ and $\{x_{it}, x_{it}^2\}$. To avoid overfitting, I set the maximum 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.

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.

References

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

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.

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

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

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

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.