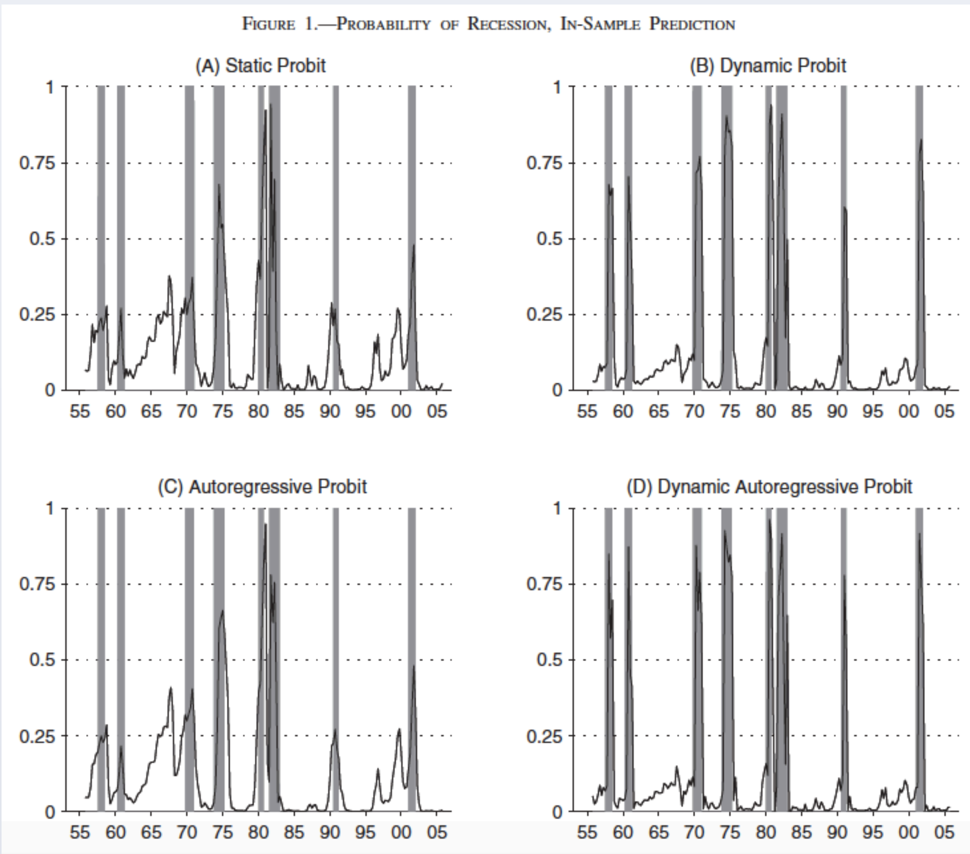
# Forecasting U.S. Recessions with a Large Number of Predictors


## Shuting Chen

### Perspectives on Computational Research (MACS30200), University of Chicago


## Introduction

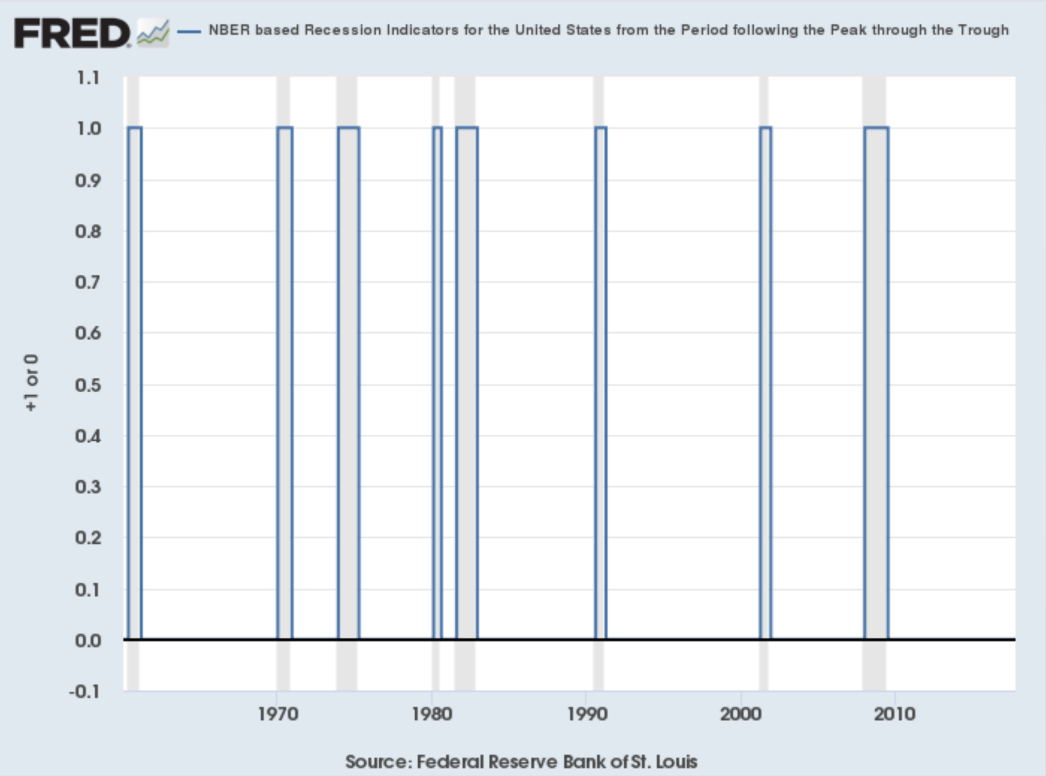
This project introduces new dynamic autoregressive probit models including factors selected by principal component analysis and predictors determined by adaptive boosting and apply them to predict recessions in the United States. The underlying motivation is based on the possibility to improve models' predictive power shown in Figure 1 (Kauppi and Saikkonen, 2008).




## Dataset

This projects uses two sources of data:

1. **U.S. business cycle expansion and contraction dates announced by NBER**




2. **Monthly frequency macroeconomic series from FRED – MD**
   - Including 123 variables over 1960:M1 to 2018:M4
   - Covering macroeconomic and financial series such as real activity indicators, interest rate indices and price indices

## Dataset (cont.)

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 |

## Methods

- **Model:** since $y_t$ (whether or not the state is in a recession) is a binary response variable, this project uses probit model $\Phi(\pi_t)$ to predict the conditional probability for a recession, where $\pi_t$ is specified as

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

- $f_{t-1}$ is a general form representing a vector of *common factors extracted by principal component analysis* or *predictors determined by adaptive boosting*. Each factor or predictor could be chosen at different lag orders greater than or equal to the forecasting horizon.
- **'Stepwise' Selection Procedure:**
- Consider first 10 principal components (or selected predictors) and allow 1-6 lag orders for each of them:
  - Initially 60 potential regressors
  - Whether or not they should be included in $\pi_t$ is determined by BIC
- **Step 1:** consider 60 models, each of which includes only one of 60 lagged factors as $f_{t-1}$ in $\pi_t$. Estimate all 60 models and pick the best one with the lowest value of BIC.
- **Step 2:** 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.
- **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.

## Methods (cont.)

- **Forecasting procedures:**
  - h = 3, 6, 12 months ahead
  - Predict the probability by iterative approach elaborated in Kauppi and Saikkonen (2008)
  - In-sample prediction: covering data from 1960:M1 to 2018:M4
  - Out-of-sample prediction:
    - 120-month estimation period
    - Rolling-window forecasting
    - Example: to predict $y_{1979:M1}$, using data from 1969:M1– h to 1979:M1 – h for $y_t$ , 1969:M1 – 2h – 5 to 1979:M1 – 2h for macro variables

## Results

**In – sample estimation (Table 2):**

Overall, $\hat{\omega}$ are all negative while $\hat{\alpha}$ are significantly positive. This is expected since the current state of an economy should have a strong positive correlation with the economic state in the last period. Comparing coefficients in the right panel to those in the left panel, we can conclude that including nonlinearity makes estimates have smaller standard errors.

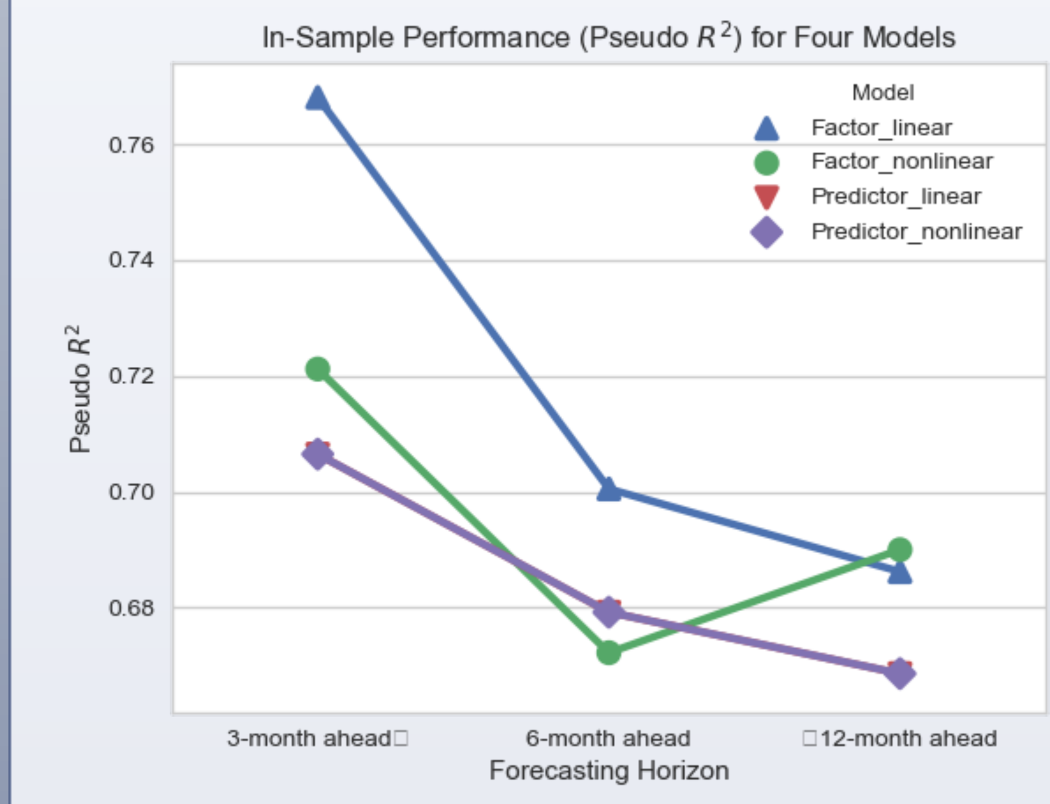
Table 2: In-sample estimation: common factors

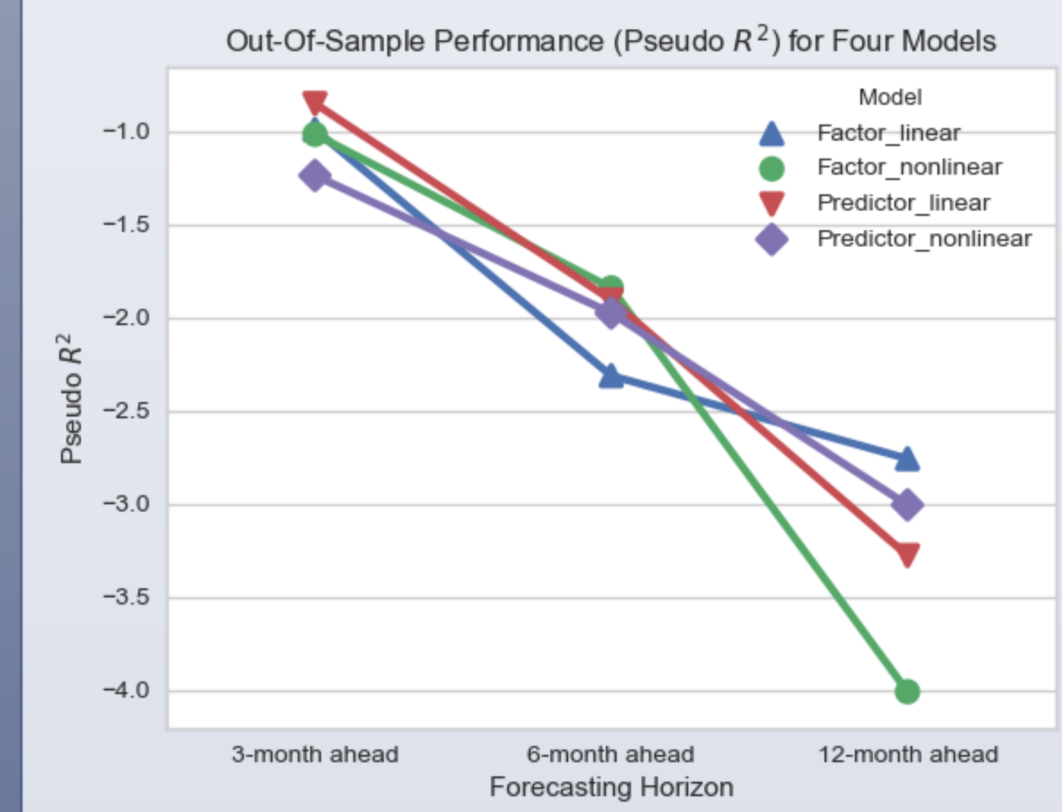
| | Factors selected from $\{x_{it}\}$ | | | Factors selected from $\{x_{it}, x_{it}^2\}$ | | |
|---|---|---|---|---|---|---|
| | $h=3$ | $h=6$ | $h=12$ | $h=3$ | $h=6$ | $h=12$ |
| $\hat{\omega}$ | -3.6224 | -3.1210 | -3.1711 | -2.3522 | -2.9032 | -3.1765 |
| | (0.5909) | (0.4664) | (0.4553) | (0.1671) | (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) |

**In – sample and out-of-sample performance:**

The following two figures display the estimated pseudo $R^2$ for four proposed models as an evaluation of in-sample and out-of-sample performance. Overall, the value of pseudo $R^2$ decreases as the forecasting horizon increases. However, all pseudo $R^2$ are negative for out-of-sample performance. This means that these models likely experience the issue of overfitting.

## Results (cont.)




Comparing the prediction performance among these four models, the models with factors selected by PCA perform better than the selected predictors. This seems to verify that the principal components with most important information would improve the prediction of recessions in the near future. Besides, there is no significant evidence that the forecasting performance would be improved by including nonlinearity of macroeconomic series.

## References


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


## Acknowledgements


I would like to express my sincere gratitude to Dr. Evans for his strong support and useful comments.