Forecasting Recessions in a Big Data Environment

by

Max Sties

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Economics
University of Alberta

Abstract

This thesis examines the predictability of Canadian recessions with special emphasis on variable selection in a big data environment. The first paper in this thesis addresses the problem of variable selection from a traditional point of view by employing a prescreened set of selected individual variables as well as data aggregation via factor analysis. Dynamic factors are estimated from panels of macroeconomic time series for Canada and the US. The factors are derived from financial, stock market, and real activity indicators for both countries. The predictive power of these factors is compared to the power of observed data. Additionally, the predictive content of US versus domestic data is evaluated. Results show that factor augmented probit regressions outperform models based solely on observed data, with a real-activity factor performing particularly well at short forecast horizons. Further, while at longer forecast horizons US interest rate spreads are consistently part of the best performing models, there is little gain in predictive accuracy from adding US data. The second paper uses modern machine learning techniques that allow for a much larger set of candidate variables. Logistic lasso and gradient boosting perform variable selection and model estimation simultaneously, thus making variable prescreening obsolete. The algorithms identify new leading indicators of recessions as well as provide evidence of structural instability in the forecasting model. I find that variables from the US labour and housing market best complement Canadian yield spreads as short term indicators, particularly during the 2008/2009 recession when yield spreads lose predictive power. Longer term forecasts are dominated by Canadian yield spreads and other financial indicators. US yield spreads and variables from the Canadian oil and gas sector do not hold predictive power at any forecast horizon.

To Anna and Lea

Acknowledgements

I am deeply indebted to my supervisor Sebastian Fossati for his continuous and unwavering support. Special gratitude also goes to him and Rodrigo Sekkel as coauthors of the first paper in this thesis. I would like to specifically thank Dmytro Hryshko and Haifang Huang for being productive and helpful members of my committee. Additional gratitude goes to Denise Young, Malik Shukayev, Valentina Galvani, as well as participants of the 2014 and 2016 Canadian Economics Association meetings for useful comments and suggestions. Any remaining errors or omissions are my sole responsibility.

Contents

Introduction **1**

1 Forecasting Recessions in Canada **4**

1.1 Introduction 4
1.2 Methodology 9
1.2.1 Dynamic Factors 9
1.2.2 Predictive Probit Regressions 11
1.2.3 Bayesian Model Averaging 15
1.3 Results 17
1.3.1 In-Sample Results 17
1.3.2 Out-of-Sample Results 22
1.4 Conclusion 26
Data Appendix 29

2 Forecasting Canadian Recessions: Making Use of Supervised Machine Learning **32**

2.1 Introduction 32
2.2 Data 35
2.3 Two Supervised Learning Methods 38
2.3.1 Lasso 38
2.3.2 Boosting 40
2.4 Model Validation 44
2.4.1 Goodness of Fit Measures 45
2.4.2 k-fold Cross Validation 46
2.4.3 k-step Ahead Time Series Cross Validation. 48
2.4.4 Cross Validation with Binary Reference Series 49
2.5 Estimation Results 51
2.5.1 Recursive Out-of-Sample Forecasting 51
2.5.2 Alternative Model Specifications 54

2.6 Variable Selection 60
2.6.1 Full Sample Variable Selection 60
2.6.2 Out-of-Sample Variable Selection 61
2.7 Conclusion 66
Appendix 68
A.1 Description of Selected Variables 68
A.2 Data Transformation Algorithm 69
A.3 Boosting Algorithm 70
A.4 More on AUC and ROC 71

Conclusion **72**

Bibliography **75**

List of Figures

1.1 Full sample estimates (posterior means) of the CA dynamic factors. Shaded areas denote recession months in Canada according to the chronology of the C.D. Howe Institute (Cross and Bergevin, 2012). . . 12

1.2 Full sample estimates (posterior means) of the US dynamic factors. Shaded areas denote recession months in Canada according to the chronology of the C.D. Howe Institute (Cross and Bergevin, 2012). . . 13

1.3 In-sample predictive content of CA and US indicators and factors. Gray lines indicate R^2_{mf} coefficients of observable predictors, blue lines indicate R^2_{mf} coefficients of the dynamic factor estimated from the corresponding group of indicators. 18

1.4 In-sample BMA weights for each of the 2625 3-variable probit models at different forecast horizons. 21

1.5 In-sample predicted probabilities of recession for the best performing 3-variable probit models at different forecast horizons: with factors (blue); without factors (gray). 22

1.6 Out-of-sample BMA weights for each of the 2625 3-variable probit models estimated recursively at different forecast horizons. 25

1.7 Out-of-sample predicted probabilities of recession for the best performing 3-variable probit models at different forecast horizons: with factors (blue); without factors (gray). 26

2.1 Canadian GDP and C.D. Howe recession months illustrated as shaded areas 37

2.2 Coefficients at different logarithmized values of the regularization parameter λ derived from full sample lasso estimation. The upper x-axis depicts associated degrees of freedom, i.e. number of variables with non-zero coefficient. Each line illustrates the coefficients of a certain variable. 39

2.3 Cross validation error using different goodness of fit measures derived from full sample lasso estimation at different values for the logarithmized regularization parameter λ. The upper x-axis depicts the associated number of variables with non-zero coefficient. The first dotted line from the left indicates the minimum CV, the second dotted line indicates the smallest CV within one standard error of the minimum CV. The blue line plots the corresponding in-sample error. 47
2.4 Distribution of the optimal number of variables obtained from repeated cross validation of full sample lasso estimation using a different number of validation folds (nfolds) and different goodness of fit measures (measure). 49
2.5 Estimated out-of-sample probabilities of the default lasso and boosting models validated by cross validation at different forecast horizons (h). Best Subset depicts out-of-sample probabilities derived from best subset selection logit regression. 52
2.6 Estimated out-of-sample probabilities using 5-fold and 10-fold cross validation with AUC and LPS as selection criterion at different forecast horizons 56
2.7 Estimated out-of-sample probabilities of the Elastic Net and Adaptive Lasso compared to the default lasso model at different forecast horizons 58
2.8 Estimated out-of-sample probabilities of stochastic boosting (with bagging fractions (BF) of 0.6 and 0.8) and pairwise boosting compared to the standard boosting model, validated via 5-fold cross validation using the AUC as model selection criterion at different forecast horizons. . . 59
2.9 Estimated out-of-sample probabilities of default boosting and lasso model, as well as adaptive lasso considering a 12 month publication lag for reference series at different forecast horizons. 59
2.10 Out-of-sample variable selection using lasso at forecast horizon $h = 1$. Darker shading illustrates coefficients of higher absolute value. 64
2.11 Out-of-sample variable selection using lasso at forecast horizon $h = 12$. Darker shading illustrates coefficients of higher absolute value. 65
2.12 ROC curve derived from in-sample lasso estimation with $h = 0$ 71

List of Tables

1.1 Model comparison of in-sample results 20
1.2 Model comparison of out-of-sample results 31

2.1 Demonstration of Boosting Algorithm 42
2.2 Variability of Cross Validation with Random Subsampling 50
2.3 Goodness of Fit of Default Models 53
2.4 Goodness of Fit Using Various Validation Methods 55
2.5 Goodness of Fit Using Alternative Model Specifications 57
2.6 Top 10 Variables Lasso 62
2.7 Top 10 Variables Boosting 62
2.8 Short Description of Selected Variables 68

Introduction

Recessions are pronounced, pervasive, and persistent declines in aggregate economic activity that are associated with high welfare costs for society. In major economies the dating of recession months is left to a commission of experts. Monthly recessions are defined by NBER for the US and CEPR for the EU. In Canada, recession months are dated by the Business Cycle Council of the C.D. Howe Institute. This recession indicator is published with considerable delay of up to a year. Forecasting this binary reference series therefore provides valuable information to business, policy makers and the public. An exploration into which data and models should be used for forecasting is the subject of this thesis.

To forecast recessions, a multitude of economic variables are publicly available that are potentially good predictors but likely have little or no individual predictive power. A common issue in the discipline of forecasting is making efficient use of such sparse and high dimensional data space. Supervised machine learning methods address such big data problems by performing variable selection and model estimation simultaneously and are therefore becoming increasingly popular in econometrics (Varian, 2014; Horowitz, 2015; Athey and Imbens, 2017).

In the recession literature, researchers traditionally face the problem of either arbitrarily selecting a small subset of variables to be included in a predictive model or to employ dimensionality reduction techniques. Following the former approach, Estrella and Mishkin (1998), Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) forecast US recessions focusing mainly on the yield spread as explanatory variable. The importance of the yield spread as leading indicator is mirrored by Atta-Mensah and Tkacz (1998) for the Canadian case. Hao and Ng (2011) extend the analysis to a set of 13 selected macro variables. A problem that arises in each of these studies is that it is not *a priori* clear which variables should be included in the forecasting model. While hundreds of macro variable are readily available, computational feasibility traditionally limits the analysis to a small set of selected or pre-screened variables.The latter approach to address this degrees of freedom problem are dimensionality reduction methods that aggregate the data space. Dynamic factor models retrieve a common autoregressive trend from a small number of selected explanatory variables (Stock and Watson, 1989, 1992); principal component analysis extracts multiple static factors from a large set of variables (Stock and Watson, 2006). Chen *et al.* (2011) and Fossati (2015) employ large sets of macro data to predict US recessions, while Gaudreault *et al.* (2003) use dynamic factors retrieved from a small set of selected variables to predict recessions in Canada.

The first paper in this thesis, "Forecasting Recessions in Canada"[1] combines these two traditional methods. We employ best subset selection and Bayesian model averaging to contrast the predictive power of models containing only Canadian data to models that add US data as well as dynamic factors. The factors are estimated from small sets of financial, real activity and stock market variables. The factor estimated from Canadian real activity variables is shown to tremendously improve short run forecasting results, while the inclusion of US variables contributes to forecasts of longer horizons. By assessing the predictive power of distinct groups of predictors, the paper provides a novel look at the problem of variable selection when forecasting recessions.

The second paper in this thesis, "Forecasting Canadian Recessions: Making use of Supervised Machine Learning", addresses the problem of model and variable selection from a more modern perspective. I employ two supervised machine learning methods,

[1]Coauthored with Sebastian Fossati (University of Alberta) and Rodrigo Sekkel (Bank of Canada.)

gradient boosting and logistic lasso, to select variables from a set of over 5000 predictors. Gradient boosting (Friedman *et al.*, 2000) has recently been employed by Ng (2014) to forecast US recessions. Penalized regression methods like lasso[2] (Tibshirani, 1996) are just starting to take a foothold in the field of econometrics. The application of logistic penalized regression to the problem of forecasting recession probabilities is new. I employ several specifications of the lasso and boosting model to forecast Canadian recessions at different forecast horizons. Each specification is evaluated with respect to its predictive accuracy in an out-of-sample forecasting exercise. The non-discriminatory approach to variable selection allows me to identify new variables that serve as important leading indicators and have previously been ignored in the literature. I further demonstrate how the optimal forecasting model changes over time, providing evidence of structural instability. The paper should also be seen as an exploration into forecasting economic variables in a big data environment, as the discussed methods can easily be applied to related problems.

[2] Lasso stands for 'least absolute shrinkage and selection operator'.

Chapter 1

Forecasting Recessions in Canada

1.1 Introduction

Predictions about the state of the economy figure prominently on the decision making process of households, firms, and policy makers. At least since Burns and Mitchell (1938), assessing the state of the economy to identify early warning signals of recession has been a prominent task in the discipline of macroeconomics. Renewed interest in the topic was sparked by Stock and Watson (1989, 1993) who construct coincident and leading indicators for the US economy and launched a new literature on forecasting the probability of recession. Stock and Watson (1989) interpret "the state of the economy" as an unobservable reference series reflecting co-movements in a broad range of macroeconomic aggregates such as output, employment, and sales. The reference series is estimated via dynamic factor analysis and used to construct a recession index (nowcasting) as well as recession probability forecasts. The Business Cycle Dating Committee of the National Bureau of Economic Research (NBER) partially relies on the Stock & Watson methodology to determine ex post recession dates for the US. The Business Cycle Council of the C.D. Howe institute determines these dates for Canada. These recession dates, usually published with considerable delay, can be characterized as a binary reference series, and estimating recession probabilities can be interpreted as a binary classification problem.

A rich body of literature examines the predictability of recessions in the US, with comparatively little work focusing on Canada. Estrella and Mishkin (1998) estimate univariate probit models on a series of 27 macro-variables and find that the yield spread, measured as the difference between 10 year and 3-months treasury yields, is the single most important predictor of US recession. Estrella and Hardouvelis (1991) further explore the predictive power of yield spreads as a measure of the yield curve, allowing for a greater amount of covariates. Mirroring these results for Canada, Atta-Mensah and Tkacz (1998) find that the Canadian yield spread, the difference between long term bond yields and the 3-month commercial paper rate, is the most useful indicator to predict recessions in Canada. Bernard and Gerlach (1998) further generalize this result, demonstrating that the predictive power of domestic yield spreads holds for all developed countries in their sample, including Canada. Additionally they find that the inclusion of the US yield spread adds predictive power at medium and long term forecasts for Canada. Subsequent work mainly focuses on advancements in methodology using mainly the yield spread as predictor of US recessions.

Building on the idea of Dueker (1997) to include lagged values of the recession series as predictor, Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) develop dynamic and autoregressive probit models. Nyberg (2010) and Ng (2012) extend these models by including a larger number of explanatory variables such as measures of perceived risk in the economy. Hao and Ng (2011) expand on the list of variables used to predict recessions in Canada by considering a small number of macroeconomic financial and real activity indicators, such as a measure of inflation, employment, monthly GDP and housing starts. The authors estimate a series of dynamic probit models, arguing that real activity variables have marginal predictive power over financial ones. They find that while dynamic probit specifications are better at predicting the duration of recessions, static versions of the model better predict turning points. However, due to limitations on computational feasibility as well as concerns about overfitting, the number of variables that can be used in direct recession probability forecasts is ultimately limited. Only recently the use of machine learning techniques has allowed direct probability forecasts with large datasets. For the US, Ng (2014) uses a gradient boosting algorithm that selects relevant variables sequentially, Fornaro (2016) imposes shrinkage on the coefficients of a probit model in a Bayesian version of ridge regression. Sties (2017) uses penalized regression and boosting to forecast Canadian recessions.

Parallel to these developments, advances were made using factor models that aggregate the data space and thus discount the problem of variable selection. Building on the methodology of Stock and Watson (1993), Chauvet (1998) and Kim and Nelson (1998) incorporate the prediction of recession probabilities directly into estimation of the dynamic factor using Markov regime switching. These models work well for real time nowcasting (Chauvet and Piger, 2008). At longer forecast horizons however, estimation becomes computationally expensive as the model's state dependency does not allow to make direct forecasts several periods ahead. With the availability of increasing amounts of macroeconomic data in the 2000's, the factor model literature split into two avenues as discussed in (Stock and Watson, 2016): (a) static factor models extracting information from large datasets (Stock and Watson, 2002) and (b) dynamic factor models extracting information from small panels of macroeconomic data or targeted predictors (Ludvigson and Ng, 2010). Chen *et al.* (2011) and Fossati (2016) employ large factor models to forecast US recessions. Fossati (2015) employs small dynamic factors retrieved from three panels of financial, stock market and real activity macro variables. For Canada, to the best of our knowledge, only Gaudreault *et al.* (2003) estimate a dynamic factor to nowcast recession probabilities. The factor is derived from total employment, real manufacturing shipments, real retail sales, total housing starts and a U.S. coincident economic index.

In this paper we assess the predictability of Canadian recessions using different sets of predictors at different forecast horizons. First, we retrieve six dynamic factors from three macroeconomic panels of financial, stock market, and real activity indicators from US and Canadian data, respectively. We then construct a dataset including the six factors as well as individual indicators from Canada and the US and compare the forecasting performance of models with and without these factors, as well as with and without US indicators and factors. In order to deal with model uncertainty, we compare the predictions of the best individual model from each subset, as well as a Bayesian model average (BMA) of all models within a certain subset. By comparing the forecast accuracy of models constrained to different subsets of predictors, we attempt to answer the following three questions:

(1) Do indicators from the US economy add significant predictive power to forecasts of Canadian economic activity? Bernard and Gerlach (1998) show that including the US yield spread as a second predictor in a probit regression of Canadian recessions on the domestic yield spread significantly improves forecast accuracy in-sample. However, it is not clear if this result holds out-of-sample or when more predictors from the Canadian economy are added to the regression. On the other hand, it could be the case that business conditions in Canada so strongly depend on business conditions in the US that predictors from the US economy overall outperform predictors from the Canadian economy. For inflation, Gosselin and Tkacz (2010) find evidence that dynamic factors estimated from US data are better predictors of Canadian inflation than factors estimated from Canadian data. We find that, although at longer forecast horizons US interest rate spreads are consistently part of the best performing models, there is little gain in predictive accuracy from adding US observed predictors and factors.

(2) Do dynamic factor models that aggregate information from various macroeconomic indicators yield better predictions than forecasts that rely solely on the underlying indicators as predictors? Castle *et al.* (2013) find that dynamic factors are better at forecasting GDP at short forecast horizons, while their relative performance declines as the forecast horizon increases. We find that this result holds for forecasts of recessions. Models augmented with dynamic factors significantly outper-

form models estimated with observed data alone. Among the estimated factors, the Canadian real activity factor is particularly successful at predicting turning points at short horizons. On the other hand, we find that financial factors estimated from interest rate spreads and exchange rates are the most successful at predicting recessions at 6 to 12 months horizons, however, such long term forecasts including factors improve only marginally upon forecasts made with the observed data alone.

(3) Does a Bayesian Model Average yield an improvement in forecast accuracy over individual models? Following Raftery (1995), Hoogerheide *et al.* (2010) find that BMA significantly reduces forecast error for US GDP growth and the S&P 500 index. Berge (2015) uses BMA to combine univariate probit models to forecast US recessions. While averaging forecasts of a continuous series like GDP or inflation makes intuitive sense as forecasts below the realized value of the reference series average out with forecasts above the realized value, the picture is less clear in the context of probability forecasts. The model with the best fit in-sample will predict the lowest probabilities of recession for non-recession months and probabilities close to one during recession months. The forecasts produced with this model are not likely to be improved by being averaged with models of worse fit. However, an average forecast might be more robust to out-of-sample estimation. Our findings indicate that the answer to this third question is unequivocally a negative one. A Bayesian average of all models does not improve upon the probability forecasts from the best individual model within the respective subset. This holds true in-sample, where the prediction error of the BMA is similar to the best individual model, but even more so out-of-sample, where the BMA forecasts perform significantly worse than the best individual models.

The remainder of this paper is organized as follows. Section 1.2 describes the data as well as the methodology used to estimate the dynamic factors and probit regressions. Section 1.3 summarizes the estimation results and section 1.4 concludes the paper.

1.2 Methodology

In this section we describe the empirical methodology used in this paper. In section 1.2.1 we discuss the estimation of dynamic factors from six subsets of Canadian and US macroeconomic indicators. Next, we use these estimated factors together with selected individual indicators to generate recession probabilities for the Canadian economy. We present the predictive probit regressions and forecast evaluation statistics used to select the best performing models in section 1.2.2. Finally, in section 1.2.3 we discuss the BMA strategy used to combine forecasts.

1.2.1 Dynamic Factors

Dynamic latent factors are estimated from 28 different indicators for the Canadian economy and 30 indicators for the US economy. For estimation, the data set is organized into six small panels or blocks.[1] For the US, we follow Fossati (2015) and consider three panels: (1) a bond and exchange rates data set of 22 financial indicators including interest rates, interest rate spreads, and exchange rates; (2) a data set of 4 stock market indicators including stock price indexes, dividend yield, and price-earnings ratio; (3) a data set of 4 real activity indicators including industrial production, personal income less transfer payments, real manufacturing trade and sales, and employment. Dynamic factors estimated from each of these three panels have been found useful in many forecasting exercises. For example, Ludvigson and Ng (2010) show that an important amount of variation in the two-year excess (US) bond returns can be predicted by factors estimated from panels (1) and (2). Likewise, the real activity variables in panel (3) have been used in Stock and Watson (1989), Diebold and Rudebusch (1996), Kim and Nelson (1998), Chauvet (1998), Chauvet and Piger (2008), Camacho *et al.* (2015), and Fossati (2015, 2016), among others, to model real-time business conditions in the US economy. For Canada we also construct three small panels of Canadian indicators with similar characteristics as those for the

[1] See Ludvigson and Ng (2010) for a more detailed motivation to organize the data into blocks.

US economy. The three panels are: (1) a bond and exchange rates data set of 19 financial indicators including interest rates, interest rate spreads, and exchange rates; (2) a data set of 5 stock market indicators including stock price indexes, dividend yield, and price-earnings ratio; (3) a data set of 4 real activity indicators including housing starts, production in manufacturing, credit card debt, and male employment. In contrast to the US literature described above, the literature on dynamic factors for the Canadian economy is small. For example, Gaudreault *et al.* (2003) and Bragoli and Modugno (2016) use real activity dynamic factors estimated using both Canadian and US data to nowcast business conditions in Canada. Similarly, Gosselin and Tkacz (2010) use dynamic factors also estimated using both Canadian and US data to successfully forecast the inflation rate in Canada.

For each of these six panels we estimate a dynamic factor model using Bayesian methods and the following framework. Let x be a $T \times N$ panel of macroeconomic indicators where x_{it}, $i = 1, \ldots, N$ and $t = 1, \ldots, T$, has a factor structure of the form

$$x_{it} = \lambda_i(L) g_t + e_{it} \tag{1.1}$$

where g_t is an unobserved dynamic factor, $\lambda_i(L) = \lambda_{i0} + \lambda_{i1} L + \cdots + \lambda_{is} L^s$ a polynomial of order s, λ_{ij} are the dynamic factor loadings, and e_{it} the idiosyncratic error. The dynamics of the latent factor and of the idiosyncratic errors are driven by autoregressive processes such that

$$\phi(L) g_t = \eta_t, \qquad \eta_t \sim i.i.d.\, N(0, \sigma_g^2) \tag{1.2}$$

$$\psi_i(L) e_{it} = \nu_{it}, \qquad \nu_{it} \sim i.i.d.\, N(0, \sigma_i^2) \tag{1.3}$$

where $\phi(L)$ and $\psi_i(L)$ are polynomials of order p_g and p_e, respectively. The factor model is specified by assuming $s = 2$ and $p_g = p_e = 1$ for every panel so that $\lambda_i(L) = \lambda_{i0} + \lambda_{i1} L + \lambda_{i2} L^2$, $\phi(L) = 1 - \phi L$, and $\psi(L) = 1 - \psi L$ for $i = 1, \ldots, N$. For estimation, the dynamic factor model is written in state-space form and estimated via Gibbs sampling following Kim and Nelson (1999) and Ludvigson and Ng (2010).[2] Identification is achieved by setting $\lambda_{10} = 1$, that is the factor loading on the first time series in each panel to 1. Finally, the parameters λ_{ij} and ψ_i are initialized to zero, σ_g^2, and σ_i^2 are initialized to 0.5, and principal components is used to initialize the dynamic factor. The Gibbs sampler runs 6,000 times. After discarding the first 1,000 draws (burn-in period), posterior means are computed using a thinning factor of 10, that is computed from every 10th draw.

The variables included in each panel, as well as their sources and the transformations employed, are described in the appendix. Our data set starts in 1967:1 and ends in 2010:12. Prior to estimation, the data are transformed to ensure stationarity and standardized. Since real activity variables are usually available with some lag, we account for data availability at time t by using the last known value x_{it-p}, where p indicates the publication lag of variable i. Publication lags for US indicators are adopted from Katayama (2010). Publication lags for Canadian real activity indicators are obtained from Statistics Canada. Figures 1.1 and 1.2 depict the three estimated factors from the full sample of Canadian and US data, respectively. The shaded areas indicate recession periods in Canada (Cross and Bergevin, 2012). Both sets of factors display similar characteristics. For example, periods of recession are coincident with dips in the real activity factors and major troughs correspond closely to Canadian recession dates. On the other hand, dips in the financial factors seem to precede recession periods. Finally, the stock market factors are characterized by higher volatility and no obvious correlation with recession months emerges from this plots.

1.2.2 Predictive Probit Regressions

The recession indicator for the Canadian economy is defined as follows. Let y_{t+h} be a binary variable which equals 1 if the month $t + h$ is subsequently declared as a

[2] While the dynamic factors can also be estimated by maximum likelihood, Gibbs sampling provides a more robust alternative for the out-of-sample recursive exercises implemented below.

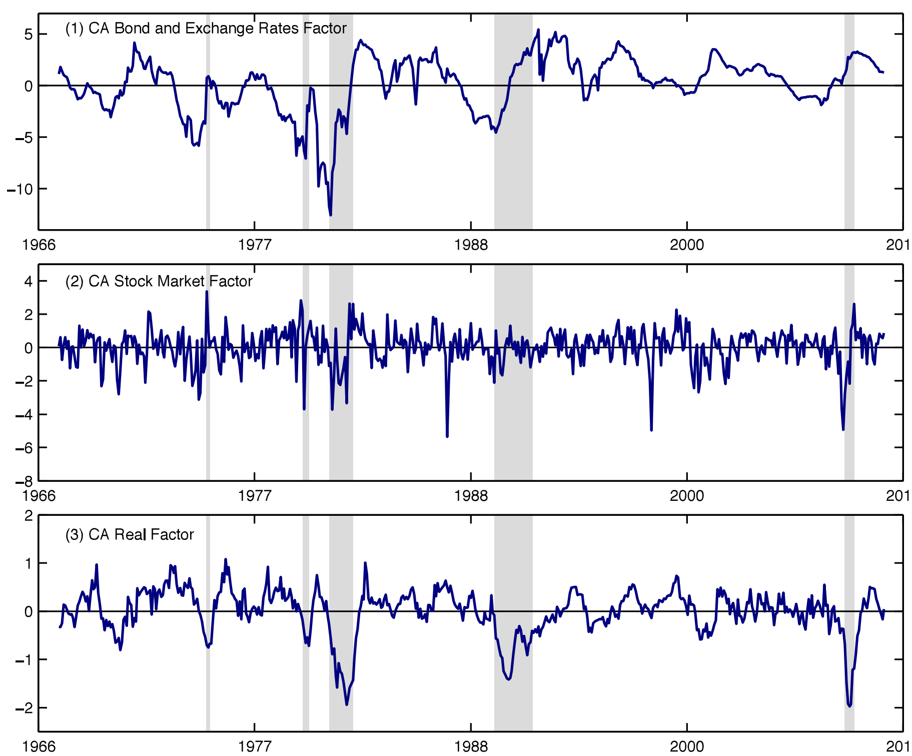

Figure 1.1: Full sample estimates (posterior means) of the CA dynamic factors. Shaded areas denote recession months in Canada according to the chronology of the C.D. Howe Institute (Cross and Bergevin, 2012).

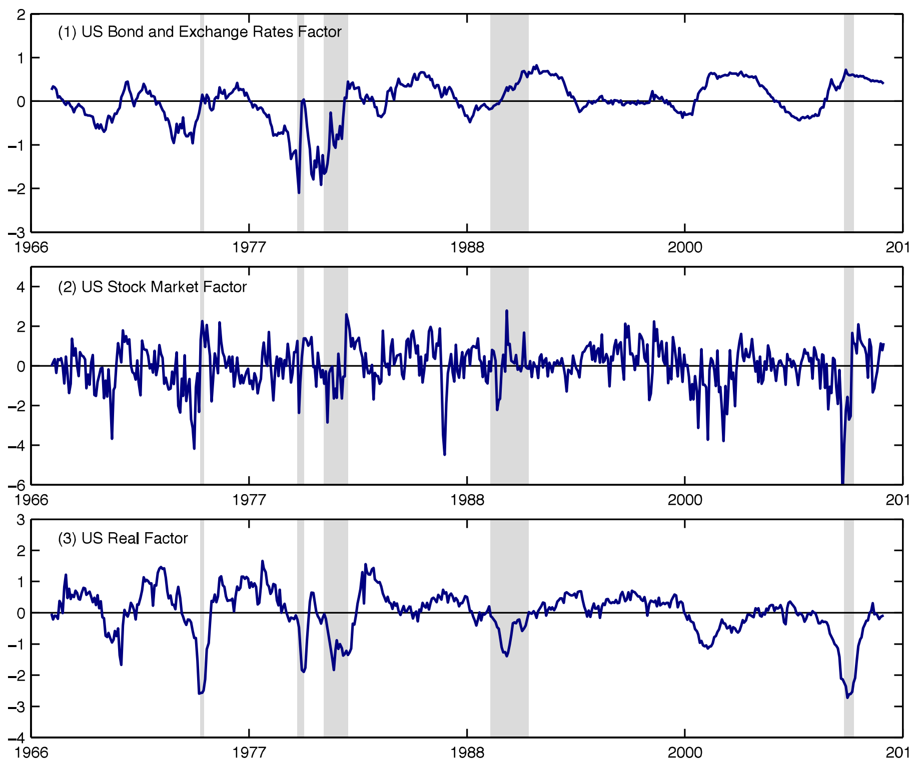

Figure 1.2: Full sample estimates (posterior means) of the US dynamic factors. Shaded areas denote recession months in Canada according to the chronology of the C.D. Howe Institute (Cross and Bergevin, 2012).

recession and 0 otherwise. A forecast of the probability of a recession in month $t+h$ (p_{t+h}) from a probit regression is then given by

$$p_{t+h} = P(y_{t+h} = 1 \mid z_t) = \Phi(\beta' z_t), \tag{1.4}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, β is a vector of coefficients, and z_t is a $k \times 1$ vector of predictors including an intercept. In this paper, we use the business cycle classification provided by Cross and Bergevin (2012) of the C.D. Howe institute.[3]

Our set of potential predictors includes 58 individual indicators and the six dynamic factors estimated from the small panels the individual indicators. To make estimation feasible, we restrict our attention to a subset of individual indicators and the six factors. The selected individual indicators (highlighted with an asterisk in the data appendix) include 14 Canadian indicators (interest rates, exchange rates, interest rate spreads, stock market indexes, and real activity variables) and 5 US indicators (interest rates, interest rate spreads, a stock market index, and industrial production). This set of 19 individual indicators is a mix of variables previously used in the literature, e.g. Hao and Ng (2011), and indicators that are found to be good individual predictors. Finally, we restrict the probit models to a maximum number of three predictors (in addition to an intercept). In total, based on the 25 predictors (including factors), 2625 models are evaluated in this best subset selection exercise.[4]

All models are estimated in-sample as well as recursively out-of-sample. We evaluate the in-sample fit of each model using McFadden's pseudo-R^2 (R^2_{mf}) which is defined as

$$R^2_{mf} = 1 - \frac{\ln \hat{L}_1}{\ln \hat{L}_0}, \tag{1.5}$$

[3] We verified the robustness of our results using the recession classification adopted by Atta-Mensah and Tkacz (1998) and Hao and Ng (2011) who extend existing series using a rule of thumb of six months of negative gross domestic product growth. Results are not significantly different to our baseline estimation using the C.D. Howe recession dates.

[4] These 2625 models include 25 one-variable models, 300 two-variable models, and 2300 three-variable models.

where $\hat{L}_1$ is the value of the log-likelihood function evaluated at the estimated parameters and $\hat{L}_0$ is the log-likelihood computed only with a constant term. Predicted probabilities of recession, both in-sample and out-of-sample, are evaluated using two popular statistics. The first statistic is the quadratic probability score (QPS), which is equivalent to the mean squared error and is defined as

$$\text{QPS} = \frac{2}{T^*}\sum_{t=1}^{T^*}(y_{t+h} - \hat{p}_{t+h})^2, \tag{1.6}$$

where T^* is the effective number of forecasts and $\hat{p}_{t+h} = \Phi(\hat{\beta}' z_t)$ is the predicted probability of recession for month $t+h$ for a given model. The QPS can take values from 0 to 2 and smaller values indicate more accurate predictions. In addition, recession probabilities are also evaluated using the log probability score (LPS) which is defined as

$$\text{LPS} = -\frac{1}{T^*}\sum_{t=1}^{T^*}\left[y_{t+h}\log(\hat{p}_{t+h}) + (1 - y_{t+h})\log(1 - \hat{p}_{t+h})\right]. \tag{1.7}$$

The LPS can take values from 0 and $+\infty$ and smaller values indicate more accurate predictions. Compared to the QPS, the LPS score penalizes large errors more heavily.

1.2.3 Bayesian Model Averaging

We use BMA to combine predicted probabilities of recession obtained from the 2625 probit regressions. One of the advantages of BMA is that its forecasts tend to improve accuracy when there is uncertainty about the true model.[5] However, there are few papers exploring BMA in the context of predicting probabilities of recession. For example, Berge (2015) uses model selection and model averaging strategies (including BMA) to evaluate the information content in many economic indicators as predictors of US business cycle turning points. Similarly, Guérin and Leiva-Leon (2014) combine recession probabilities obtained from univariate and multivariate regime-switching models using BMA and other averaging strategies. Both papers find that BMA can

[5] See, for example, Faust *et al.* (1996), Wright (2008), and Groen *et al.* (2013), among others.

yield improvements in forecast accuracy and highlight the importance of allowing for time variation in the models' weights as the best forecasting models typically change over time. In addition, we use the weights assigned to the BMA forecasts to evaluate the predictive content of the dynamic factors vis-a-vis the individual predictors.

The approach we follow to average recession probabilities is similar to Berge (2015). First, from each of the M models estimated in section 1.2.2 we obtain a forecast $\hat{p}_{t+h}$, resulting in $\{\hat{p}^1_{t+h}, \hat{p}^2_{t+h}, \dots, \hat{p}^M_{t+h}\}$. The BMA combined forecast assigns each of the M models a weight w_i, $i = 1, \dots, M$, such that

$$\hat{p}^{BMA}_{t+h} = \sum_{i=1}^{M} \hat{p}^i_{t+h} w_i \tag{1.8}$$

where $w_i = P(M_i \,|\, D)$ is the posterior probability of model i conditional on observed data D. The posterior probability of model i is given by

$$P(M_i \,|\, D) = \frac{P(D \,|\, M_i) P(M_i)}{\sum_{j=1}^{M} P(D \,|\, M_j) P(M_j)} \tag{1.9}$$

where $P(D \,|\, M_i)$ is the marginal likelihood of model i and $P(M_i)$ is the prior probability that model i is true. Calculating the marginal likelihood can be a high-dimensional and intractable problem. We follow much of the literature and use the BIC approximation as discussed in Raftery (1995). When each model is deemed to be equally likely a priori, the i-th model posterior probability can be approximated by its fit relative to the fit of all other models such that

$$P(M_i \,|\, D) = \frac{\exp(\text{BIC}_i)}{\sum_{i=1}^{M} \exp(\text{BIC}_i)}. \tag{1.10}$$

As suggested in Raftery (1995), the BIC for model i is defined as

$$\text{BIC}_i = -LR_i + k \ln T \tag{1.11}$$

where LR_i is the likelihood ratio test statistic for testing model i against a model with only a constant term, k is the number of predictors, and T is the sample size.

1.3 Results

In this section we compare the predictive performance of the different models, as well as the BMA predictions. In addition, we compare the predictive performance of models that include dynamic factors to models that do not include factors. Similarly, we compare the performance of models that include US data (factors and indicators) to models that include only Canadian data. All models are estimated in-sample (using the full set of available observations), as well as out-of-sample (using only observations up to the time the forecast would have been made to mimic real-time forecasting).

1.3.1 In-Sample Results

We start by assessing the individual in-sample predictive content of each variable at different forecast horizons. To this end, we estimate one-variable probit models by regressing the recession series y_{t+h} for $h \in \{1, \ldots, 18\}$ on each indicator and estimated factor separately. The models are estimated using data starting in 1967:3 and ending in 2010:12, that is, the full sample. Figure 1.3 plots the regression R^2_{mf} coefficients versus the forecast horizon h. Gray lines represent the coefficients for the individual indicators while blue lines depict the R^2_{mf} coefficients for the estimated dynamic factors. We present the results following the six panels described above, with models estimated with Canadian data on the left panels and models estimated with US data on the right panels. When comparing the R^2_{mf} coefficients across different forecast horizons we see similar results for US and Canadian indicators and factors, but US variables overall have lower individual predictive content for Canadian recessions. The results show that the predictive content of bond and exchange rate variables is relatively low at short horizons, but rises as the forecast horizons increase to peak at about 12-months ahead forecasts. The stock market indicators have relatively little predictive content. The average predictive content of stock market variables peaks between 3 to 6 months. On the other hand, the real activity indicators have very

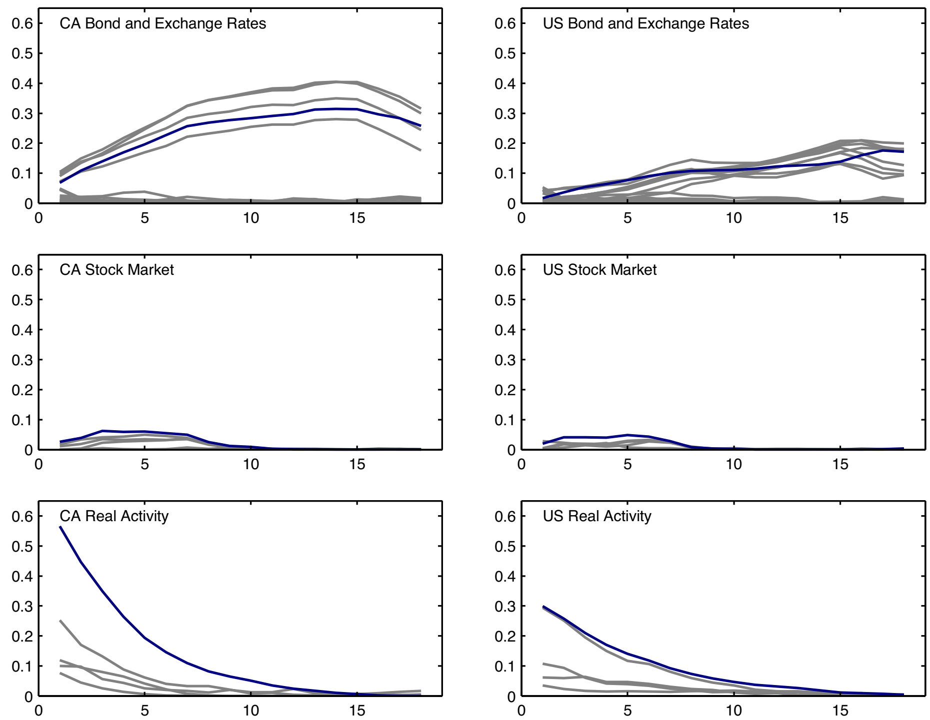

Figure 1.3: In-sample predictive content of CA and US indicators and factors. Gray lines indicate R^2_{mf} coefficients of observable predictors, blue lines indicate R^2_{mf} coefficients of the dynamic factor estimated from the corresponding group of indicators.

strong predictive content at short forecast horizons. In particular, notice that the proposed Canadian real activity factor improves significantly upon any of the R^2_{mf} coefficients for the observable indicators.

Next, we estimate all combinations of 3-variable probit models as described in section 1.2.2. Table 1.1 reports the in-sample QPS and LPS for the best individual models, as well as the BMA results. Three sets of models can be distinguished. The first set uses the observable indicators and estimated factors, and uses both Canadian and US data (first column). The second set includes models only estimated with the observable indicators, that is, without any of the estimated factors (second column). The third set includes all the models estimated using only Canadian data, that is, without any US data (third column). All variables are for the Canadian economy unless indicated otherwise. For $h = 1$, the shortest forecast horizon considered, the best model (985) includes housing starts (HS), the 10-year Canadian yield spread (YS_{10}), and the Canadian real activity factor ($real^{CA}$). If factors are excluded from the set of predictors, the best model (2312) uses credit card debt (CCD), male employment (EMP), and the 5-year yield spread (YS_5). Excluding the Canadian real activity factor, however, results in a substantial deterioration in fit (larger QPS and LPS values). Finally, since the best model includes only Canadian data, for $h = 1$ the selected model does not change when US variables are excluded.

When the forecasting horizon is increased to 3 months ($h = 3$) we find that the best performing models are the same as for $h = 1$. The models, however, exhibit a small deterioration in fit due to the reduced predictive content of the Canadian real activity factor at longer horizons. In contrast, as the forecasting horizon is increased to 6 and 12 months, the best performing models change and US variables start appearing in them. For example, at $h = 6$ the best performing model (1563) now includes the 10-year Canadian yield spread (YS_{10}), the equivalent 10-year US yield spread (YS_{10}^{US}), and the US real activity factor ($real^{US}$). For $h = 12$, the best performing model (2256) drops the US real activity factor and incorporates the 5-year US yield spread (YS_5^{US}). As a result, for 6 and 12 month ahead forecasts, the performance of the best models deteriorates when US variables are excluded from the potential set of predictors. On the other hand, when the dynamic factors are excluded we observe a deterioration in the forecasting performance at 1, 3, and 6 months, but not at 12 months.

Next, we focus on the in-sample performance of the BMA forecasts. The results reported in Table 1.1 show that BMA delivers an in-sample performance that is essentially identical to the one reported for the best performing models. For $h = 1$, Figure 1.4 shows that about 70% weight is given to the best performing model (985), while 12% weight is given to a slightly different model where the 10-year yield spread (YS_{10}) is substituted with the 5-year spread (YS_5). For $h = 3$, the same two models receive about 60% and 19% weight, respectively. Similarly, for $h = 6$ the best

Table 1.1: Model comparison of in-sample results

$h=1$	all variables		w/o factors		w/o US variables	
	Best	BMA	Best	BMA	Best	BMA
Model	985	–	2312	–	985	–
Var1	$real^{CA}$	–	YS_5	–	$real^{CA}$	–
Var2	YS_{10}	–	CCD	–	YS_{10}	–
Var3	HS	–	EMP	–	HS	–
QPS	0.07	0.07	0.11	0.11	0.07	0.07
LPS	0.11	0.11	0.19	0.19	0.11	0.11
T	525	525	525	525	525	525
$h=3$	all variables		w/o factors		w/o US variables	
	Best	BMA	Best	BMA	Best	BMA
Model	985	–	2312	–	985	–
Var1	$real^{CA}$	–	YS_5	–	$real^{CA}$	–
Var2	YS_{10}	–	CCD	–	YS_{10}	–
Var3	HS	–	EMP	–	HS	–
QPS	0.09	0.09	0.12	0.12	0.09	0.09
LPS	0.15	0.15	0.21	0.21	0.15	0.15
T	523	523	523	523	523	523
$h=6$	all variables		w/o factors		w/o US variables	
	Best	BMA	Best	BMA	Best	BMA
Model	1563	–	2316	–	985	–
Var1	$real^{US}$	–	YS_5	–	$real^{CA}$	–
Var2	YS_{10}	–	CCD	–	YS_{10}	–
Var3	YS_{10}^{US}	–	YS_{10}^{US}	–	HS	–
QPS	0.11	0.11	0.12	0.12	0.11	0.11
LPS	0.17	0.17	0.20	0.20	0.19	0.19
T	520	520	520	520	520	520
$h=12$	all variables		w/o factors		w/o US variables	
	Best	BMA	Best	BMA	Best	BMA
Model	2256	–	2256	–	1860	–
Var1	YS_{10}	–	YS_{10}	–	IBR	–
Var2	YS_5^{US}	–	YS_5^{US}	–	YS_5	–
Var3	YS_{10}^{US}	–	YS_{10}^{US}	–	CCD	–
QPS	0.10	0.10	0.10	0.10	0.11	0.12
LPS	0.19	0.19	0.19	0.19	0.20	0.20
T	514	514	514	514	514	514

Notes: The column "Best" refers to the best performing individual model according to QPS. Variable descriptions can be found in the data appendix.

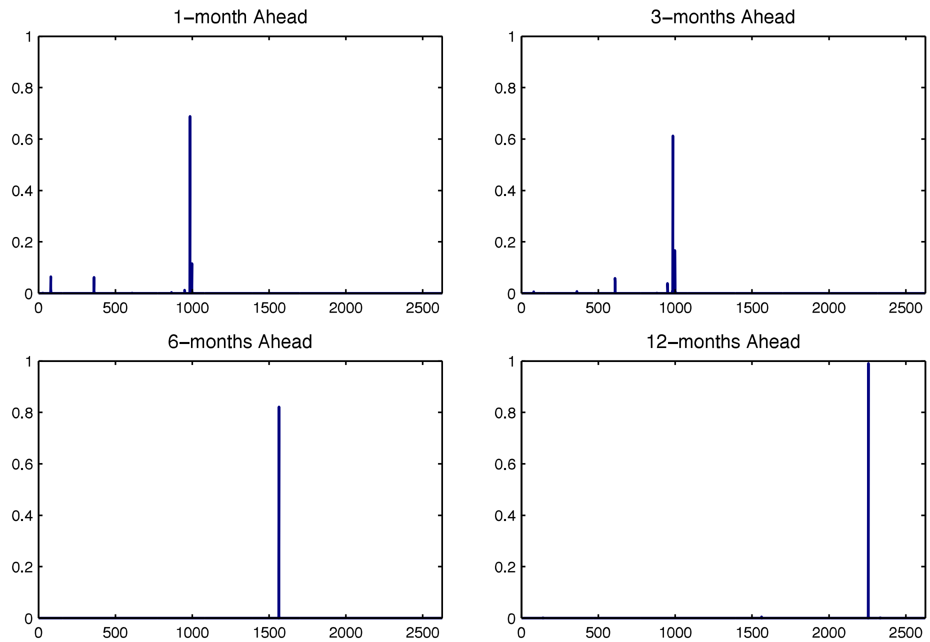

Figure 1.4: In-sample BMA weights for each of the 2625 3-variable probit models at different forecast horizons.

performing model (1563) receives 82% of the weight, while 18% weight is given to a very similar model. Finally, for $h = 12$ we find that 99% of the weight is given to the best performing model (2256). Several conclusions can be drawn from the in-sample BMA results. First, at all horizons, the best performing model according to QPS receives the highest weight in the BMA forecast. Second, BMA gives positive weight to few models, and these models are generally very similar with respect to the variables they contain. As a result, BMA weights are highly concentrated on few very effective predictors and BMA forecasts end up being very similar to the ones obtained from the best forecasting models.

In sum, our in-sample results show that Canadian real activity indicators (housing starts and employment) and particularly the Canadian real activity factor are the preferred variables for generating short term (1 to 3 months) recession probabilities of the Canadian economy. At longer horizons (6 to 12 months), the preferred variables include Canadian and US yield spreads, mainly the 10-year yield spread. In terms

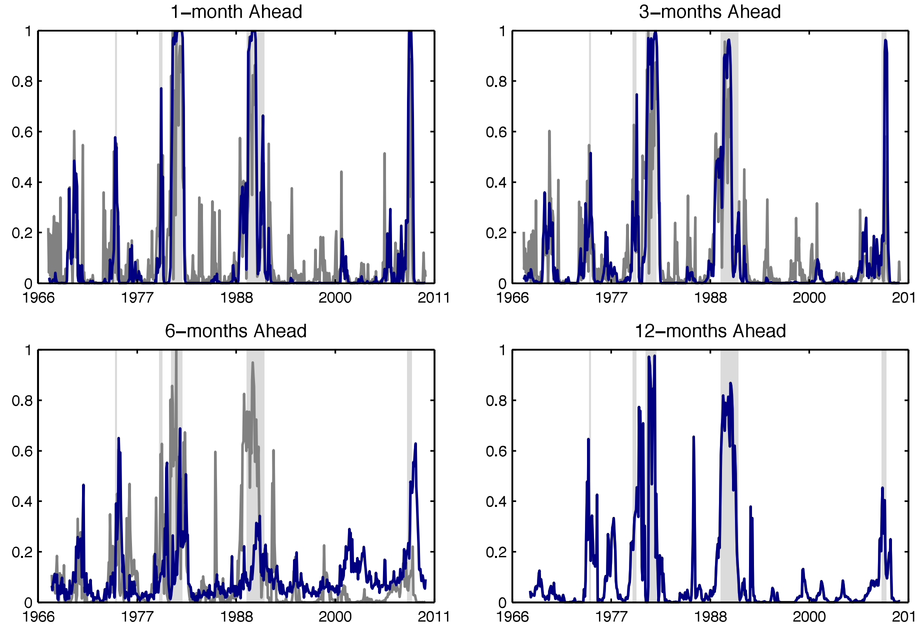

Figure 1.5: In-sample predicted probabilities of recession for the best performing 3-variable probit models at different forecast horizons: with factors (blue); without factors (gray).

of the questions formulated in the introduction, we find the following results: (1) Excluding US data results in a very small deterioration in fit at longer horizons; (2) Excluding factors can result in a substantial deterioration in fit at shorter horizons; (3) BMA forecasts cannot improve the performance of the best model selected by QPS. To illustrate point (2), Figure 1.5 shows the in-sample predicted probabilities of recession for the best performing models with and without factors. For forecast horizons of 1 and 3 months, the best model with factors produces recession probabilities that are closer to 0 during expansions and closer to 1 during recessions. This improvement, however, vanishes as the forecasts horizon is extended to 6 and 12 months.

1.3.2 Out-of-Sample Results

We now evaluate the performance of the models in a recursive out-of-sample forecasting exercise.[6] In this case, the set of observations is divided into an initial estimation

[6] This exercise uses ex-post revised data (instead of real-time data) to generate out-of-sample predicted recession probabilities for each of the models.

sample from 1967:3 to 1988:1 ($251-h$ effective observations) and a hold-out sample with the remaining observations. A direct h-step ahead forecast is produced for each period in the hold-out sample, with the first forecast made for 1988:1+h and the last for 2010:12. As a result, the hold-out sample includes 275 out-of-sample predictions when $h=1$, 273 predictions when $h=3$, 270 predictions when $h=6$, and 264 predictions when $h=12$. First, the dynamic factors are estimated recursively, each period using data available at time t, and expanding the estimation window by one observation each month. Next, the probit models are also estimated recursively and used to generate a recession probability for month $t+h$ based the information available at month t. We account for data availability at each point in time by adjusting for the publication lag in real activity variables (see, for example, Katayama, 2010; Fossati, 2015).

Table 1.2 reports the out-of-sample QPS and LPS for the best individual models, as well as the BMA results. For $h=1$, the best model (985) is the same model found in-sample and includes housing starts (HS), the 10-year Canadian yield spread (YS_{10}), and the Canadian real activity factor ($real^{CA}$). At longer horizons, the observations made in-sample largely translate to the out-of-sample results but with some differences. For example, US variables now appear more often and at shorter forecast horizons. The 10-year US yield spread (YS_{10}^{US}) is selected at $h=3$ and 6, the US real activity factor ($real^{US}$) is selected at $h=3$, the US stock market factor ($stock^{US}$) is selected at $h=6$, and the Federal Funds Rate (FF^{US}) is selected at $h=12$. But while US variables appear to be more relevant, excluding US variables from the potential set of predictors has almost no effect in the out-of-sample performance of the models (mainly larger LPS values at $h=3$ and 6). On the other hand, at shorter forecast horizons we find that factors improve the out-of-sample performance of the models and excluding the estimated factors from the set of predictors results in a deterioration in fit.

"Best" refers to the best performing individual model according to QPS. Variable descriptions can be found in the data appendix.

We now focus on the out-of-sample performance of the BMA forecasts. One advantage of averaging is that BMA weights are re-computed for each period in the hold-out sample. As a result, the models and therefore the variables that are good predictors are allowed to change over time. Figure 1.6 shows each of the model's weight for each of the out-of-sample predictions. For example, for the 1-month forecast, before the first recession in the hold-out sample the dominant model in the BMA forecast (green line) includes the Canadian real activity factor ($real^{CA}$), the US financial factor ($bond^{US}$), and the US real activity factor ($real^{US}$). During the two recession periods BMA assigns most weight to a model (blue line) that includes the Canadian real activity factor, the Canadian financial factor ($bond^{CA}$), and housing starts (HS). In contrast, between the two recession periods the dominant model (purple line) includes the Canadian real activity factor, the US financial factor, and housing starts. Finally, after the last recession BMA assigns most weight to a model that includes the Canadian real activity factor, the 10-year yield spread, and housing starts (that is, the model with best in-sample fit). The out-of-sample BMA weights for other forecast horizons paint a similar picture. For the 3-month forecasts, BMA allocates most weight to the same variables as 1-month ahead. For the 6-month forecasts, the models with highest weight include the Canadian real activity factor, the 10-year yield spread, and housing starts as the most selected predictors. Finally, the 12-month ahead forecasts include variables such as the Canadian and US financial factors, as well as Canadian and US yield spreads. At each horizon, all other models get very low weight throughout the entire hold-out sample period. As a result, although the dominant model changes over time, essentially the same set of variables is selected consistently for each forecast horizon.

In terms of the recursive out-of-sample performance of the BMA forecasts, the results reported in Table 1.2 show that averaging cannot improve the accuracy of the best models selected by QPS. In fact, averaging can result in a substantial deteriora-

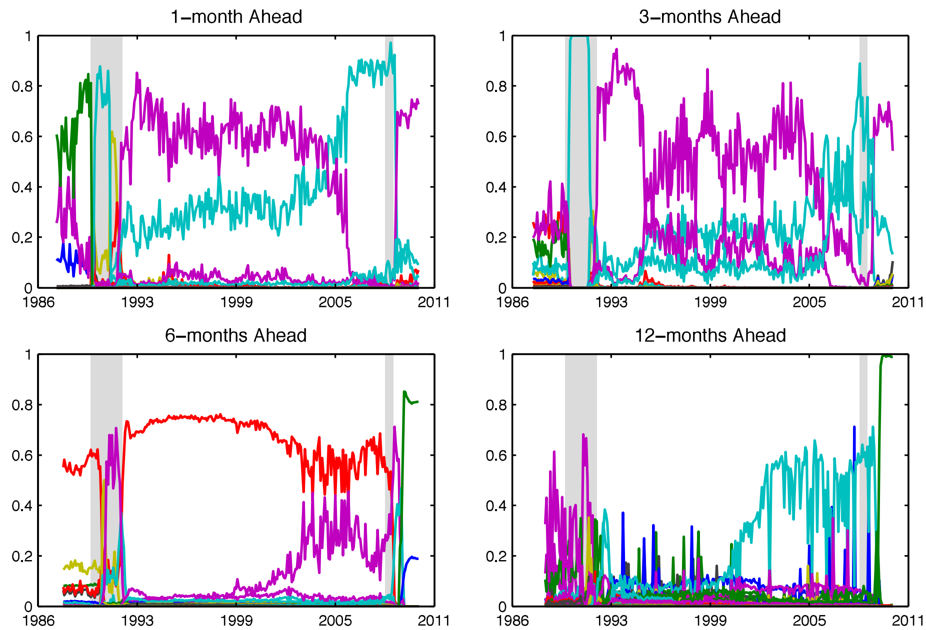

Figure 1.6: Out-of-sample BMA weights for each of the 2625 3-variable probit models estimated recursively at different forecast horizons.

tion in accuracy at longer horizons. Overall, the out-of-sample results are consistent with the in-sample results discussed above and show that real activity variables (the Canadian and US real activity factors, housing starts, etc.) are the preferred variables for generating short term (1 to 3 months) recession probabilities of the Canadian economy. At longer horizons (6 to 12 months), the preferred variables include the Canadian and US financial factors, as well as yield spreads. In terms of the questions formulated in this paper, we find the following results: (1) Excluding US data results in no substantial deterioration in out-of-sample fit; (2) Excluding factors can result in a deterioration in fit at shorter horizons; (3) BMA forecasts cannot improve the performance of the best model selected by QPS. Finally, Figure 1.7 shows the out-of-sample predicted probabilities of recession for the best performing models with and without factors. For forecast horizons of 1 and 3 months, the best model with factors produces produces recession probabilities that are closer to 0 during expansions and closer to 1 during recessions. On the other hand, no improvements are observed for

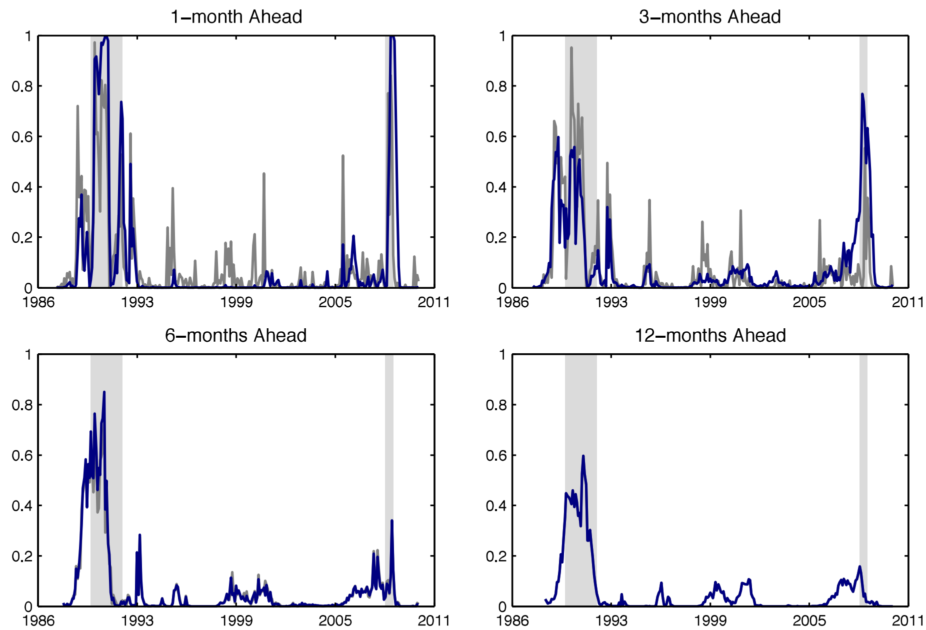

Figure 1.7: Out-of-sample predicted probabilities of recession for the best performing 3-variable probit models at different forecast horizons: with factors (blue); without factors (gray).

forecast horizons of 6 and 12 months.

1.4 Conclusion

In this paper, we show which groups of predictors can be used to best forecast recessions in Canada. We use best subset selection as well as Bayesian model averaging to compare the predictive power of models with and without dynamic factors as well as with and without US variables. We close this paper with a few concluding remarks.

Firstly, Our findings confirm the importance of domestic yield spreads in making predictions at any forecast horizon (Atta-Mensah and Tkacz, 1998). Yield spreads appear in some form in every single model selected by best subset selection and are best complemented with real activity indicators at short forecast horizons and with financial indicators at long forecast horizons. Stock market indicators generally do not exhibit significant predictive content at any forecast horizon. This mirrors the results of Fossati (2015) who finds that among three estimated factors (financial, stock market, real activity), the stock market factor holds the least predictive power to forecast US recessions.

The use of US data in forecasting Canadian recessions can improve long term forecasts but is less useful at shorter forecast horizons. This is in line with the notion that spillovers from the US affect economic conditions in Canada with a delay (Beaton *et al.*, 2014). Our findings are similar to the results of Bragoli and Modugno (2016) who find that US variables matter when nowcasting Canadian GDP and confirm the finding of Bernard and Gerlach (1998) that US yield spreads add predictive power to Canadian recession forecasts at medium and long forecast horizons, but not at shorter ones. Our results however do not go as far as Gosselin and Tkacz (2010) who find that Canadian inflation can be forecast with dynamic factors solely estimated from US data. Canadian macro indicators remain the most important predictors of Canadian recessions.

We estimate a new Canadian real activity factor that can be used to accurately predict recessions in the short term and significantly improves upon the predictive power of its underlying macro series. Our findings are in line with Castle *et al.* (2013) who show that dynamic factors perform better than observable data at short forecast horizons. Our factor, extracting the co-movement in Canadian housing starts, production in manufacturing, credit card debt and male employment, can be used as a coincident indicator due to its strong correlation with the Canadian business cycle. It can be considered an update to the factor estimated by Gaudreault *et al.* (2003). Small dynamic factors generally are better predictors than their underlying observed series due to a reduction in noise and the addition of autoregressive terms within the factor. By augmenting probit models with dynamic factors, the advantages of static probit models to predict turning points and the advantages of dynamic structures in predicting recession duration as found in Hao and Ng (2011) can be combined. Dynamic factors are therefore more robust to out-of-sample estimation.

Finally, Bayesian model averaging assigns significant weight only to few but similar models. The best individual models receive the highest weight within the in-sample BMA forecast. Out of sample, BMA forecasts perform significantly worse than the best individual models.

Data Appendix

The following tables list the short name, transformation applied, and a data description of each series in the six groups considered. Canadian data are retrieved from the statistics Canada CANSIM database as well as the OECD. All US bond, exchange rates, and stock market series are from FRED (St. Louis Fed), unless the source is listed as GFD (Global Financial Data), or AC (author's calculation). Data for the US real activity factor are from Camacho *et al.* (2015). The transformation codes are: 1 = no transformation; 2 = first difference; 3 = first difference of logarithms.

CA Variables

	Short Name	**Trans.**	**Description**
Bond and Exchange Rates Factor			
1*	BR	2	Bank rate (Percent)
2	GB.10Y	2	Governement marketable bonds average yield (over 10 years)
3	GB.5Y	2	Governement marketable bonds average yield(5-10 years)
4	GB.3Y	2	Governement marketable bonds average yield (3-5 years)
5	GB.1Y	2	Governement marketable bonds average yield (1-3 years)
6	PCP.3M	2	3 months prime corporate paper
7	PCP.2M	2	2 month prime corporate paper
8	PCP.1M	2	1 month prime corporate paper
9	MLR.5Y	2	Average residential mortgage lending rate: 5 year
10*	IBR.3M	2	Short-term interest rates; Per cent per annum
11*	EX.US	2	United States dollar, noon spot rate, average
12	EX.JAP	2	Japanese yen, noon spot rate, average
13	EX.SWIT	2	Swiss franc, noon spot rate, average
14*	EX.UK	2	United Kingdom pound sterling, noon spot rate, average
15*	YS.10Y.3M	1	Yield Spread b/t 10-yr bond and 3-m prime (AC)
16*	YS.5Y.3M	1	Yield Spread b/t 5-10-yr bond and 3-m prime (AC)
17	YS.3Y.3M	1	Yield Spread b/t 3-5-yr bond and 3-m prime (AC)
18	YS.1Y.3M	1	Yield Spread b/t 1-3-yr bond and 3-m prime (AC)
19*	M1.2005	3	Narrow Money (M1) Index 2005=100; SA
Stock Market Factor			
20*	TCI.C	3	TSX Composite Index; Close (2000=1000)
21	ST.EX.C	2	Exchange;stockyields(composite);closingquotations(Percent)
22*	SP	3	Share Prices; Index 2005=100
23	TSX.VAL	3	Toronto Stock Exchange, value of shares traded (x 1,000,000)
24	TSX.VOL	3	Toronto Stock Exchange, volume of shares traded (shares x 1,000,000)
Real Factor			
25*	HS	1	Housing starts index; total units
26*	PROD.MAN	3	Production in total manufacturing sa; 2005=100
27*	CCD	3	Credit Card Debt; At month-end; sa ; Total outstanding balances
28*	EMP.M	2	Employed population; Aged 15 and over; Males
29*	GDP	3	Gross domestic product, factor cost 1992 const. prices

US Variables

	Short Name	Trans.	Description
Bond and Exchange Rates Factor			
1*	Fed Funds	2	Interest Rate: Federal Funds (Effective) (% per annum)
2	Comm paper	2	Commercial Paper Rate
3	3-m T-bill	2	Interest Rate: USTreasury Bills, Sec Mkt, 3-Mo. (% per annum)
4	6-m T-bill	2	Interest Rate: USTreasury Bills, Sec Mkt, 6-Mo. (% per annum)
5	1-y T-bond	2	Interest Rate: USTreasury Const Maturities, 1-Yr. (% per annum)
6	5-y T-bond	2	Interest Rate: USTreasury Const Maturities, 5-Yr. (% per annum)
7	10-y T-bond	2	Interest Rate: USTreasury Const Maturities, 10-Yr. (% per annum)
8	AAA bond	2	Bond Yield: Moody's AAA Corporate (% per annum) (GFD)
9	BAA bond	2	Bond Yield: Moody's BAA Corporate (% per annum) (GFD)
10	CP spread	1	Comm paper – Fed Funds (AC)
11	3-m spread	1	3-m T-bill – Fed Funds (AC)
12	6-m spread	1	6-m T-bill – Fed Funds (AC)
13	1-y spread	1	1-y T-bond – Fed Funds (AC)
14*	5-y spread	1	5-y T-bond – Fed Funds (AC)
15*	10-y spread	1	10-y T-bond – Fed Funds (AC)
16	AAA spread	1	AAA bond – Fed Funds (AC)
17	BAA spread	1	BAA bond – Fed Funds (AC)
18	Ex rate: index	3	Exchange Rate Index (Index No.) (GFD)
19	Ex rate: Swit	3	Foreign Exchange Rate: Switzerland (Swiss Franc per US$)
20	Ex rate: Jap	3	Foreign Exchange Rate: Japan (Yen per US$)
21	Ex rate: U.K.	3	Foreign Exchange Rate: United Kingdom (Cents per Pound)
22	Ex rate: Can	3	Foreign Exchange Rate: Canada (Canadian$ per US$)
Stock Market Factor			
23*	S&P 500	3	S&P's Common Stock Price Index: Composite (1941-43=10) (GFD)
24	S&P indst	3	S&P's Common Stock Price Index: Industrials (1941-43=10) (GFD)
25	S&P div yield	3	S&P's Composite Common Stock: Dividend Yield (% per annum) (GFD)
26	S&P PE ratio	3	S&P's Composite Common Stock: Price-Earnings Ratio (%) (GFD)
Real Factor			
27*	IP	3	Industrial Production Index - Total Index
28	PILT	3	Personal Income Less Transfer Payments
29	MTS	3	Manufacturing and Trade Sales
30	Emp: total	3	Employees On Nonfarm Payrolls: Total Private

Table 1.2: Model comparison of out-of-sample results

	all variables		w/o factors		w/o US variables	
$h = 1$	Best	BMA	Best	BMA	Best	BMA
Model	985	–	2312	–	985	–
Var1	$real^{CA}$	–	YS_5	–	$real^{CA}$	–
Var2	YS_{10}	–	CCD	–	YS_{10}	–
Var3	HS	–	EMP	–	HS	–
QPS	0.11	0.11	0.13	0.14	0.11	0.11
LPS	0.21	0.26	0.24	0.26	0.21	0.25
T	275	275	275	275	275	275
	all variables		w/o factors		w/o US variables	
$h = 3$	Best	BMA	Best	BMA	Best	BMA
Model	1563	–	2312	–	952	–
Var1	$real^{US}$	–	YS_5	–	$real^{CA}$	–
Var2	YS_{10}	–	CCD	–	EX_{US}	–
Var3	YS_{10}^{US}	–	EMP	–	YS_5	–
QPS	0.15	0.18	0.16	0.16	0.15	0.18
LPS	0.26	0.45	0.28	0.30	0.29	0.42
T	273	273	273	273	273	273
	all variables		w/o factors		w/o US variables	
$h = 6$	Best	BMA	Best	BMA	Best	BMA
Model	1392	–	2259	–	986	–
Var1	$stock^{US}$	–	YS_{10}	–	$real^{CA}$	–
Var2	YS_{10}	–	YS_{10}^{US}	–	YS_{10}	–
Var3	YS_{10}^{US}	–	SP^{US}	–	MAN	–
QPS	0.15	0.18	0.16	0.17	0.16	0.18
LPS	0.31	0.40	0.30	0.36	0.35	0.40
T	270	270	270	270	270	270
	all variables		w/o factors		w/o US variables	
$h = 12$	Best	BMA	Best	BMA	Best	BMA
Model	1850	–	1850	–	1847	–
Var1	IBR	–	IBR	–	IBR	–
Var2	YS_{10}	–	YS_{10}	–	YS_{10}	–
Var3	FF^{US}	–	FF^{US}	–	CCD	–
QPS	0.13	0.21	0.13	0.14	0.13	0.21
LPS	0.21	0.37	0.21	0.22	0.21	0.35
T	264	264	264	264	264	264

Notes: The column "Best" refers to the best performing individual model according to QPS. Variable descriptions can be found in the data appendix.

Chapter 2

Forecasting Canadian Recessions: Making Use of Supervised Machine Learning

2.1 Introduction

A recession is a pronounced, pervasive, and persistent decline in aggregate economic activity that is associated with high welfare costs for society. Predicting recession probabilities through econometric modeling can therefore provide valuable information to policy makers, business and the public. To forecast recessions, a multitude of economic variables are publicly available that are potentially good predictors, many of which have little or no predictive power. A common issue in the discipline of forecasting is making efficient use of such sparse and high dimensional data space. Especially in time series data, where the number of observations is usually small compared to the number of potential predictors, degrees of freedom decrease rapidly and conventional econometric modeling techniques quickly reach their limits - the curse of dimensionality. Supervised machine learning methods address this problem by performing variable selection and model estimation simultaneously and are therefore becoming increasingly popular in econometrics (Varian, 2014; Horowitz, 2015; Athey and Imbens, 2017). The analysis in this paper is the first application of supervised machine learning to forecast Canadian recessions.

In big data environments, forecasters traditionally face the problem of either arbitrarily selecting a small subset of variables to be included in a predictive model or to employ dimensionality reduction techniques. Following the former approach, Estrella and Mishkin (1998) employ univariate static probit regression to examine the predictive power of a small set of hand selected macroeconomic variables. They find strong evidence that the yield spread, defined as the difference between 10-year and 3-months treasury bond yield, is the single most powerful predictor of US recessions. Recently this result has been questioned by Ng and Wright (2013) who argue that yield spreads have lost their predictive power during the 2008/2009 recession in the US. Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) extend the analysis allowing for dynamic and autoregressive structure in the probit specification, but focus solely on the yield spread as explanatory variable. The importance of the yield spread as leading indicator is mirrored by Atta-Mensah and Tkacz (1998) for Canada. The authors define Canadian yield spreads as the difference between the yields of long term government bonds and the 90-day commercial paper rate.

Making use of dynamic and autoregressive probit specifications, Hao and Ng (2011) expand the number of possible predictors to a set of 13 selected macro series. From this set of potential predictors they derive the best 3 variable probit model via best subset selection. Their best performing models include lagged values of the 10-year Canadian yield spread, housing starts, the growth rate of the M1 money stock and a leading economic indicator that Statistics Canada has since ceased to publish. In line with Bernard and Gerlach (1998) they find that US yield spreads hold predictive power at longer forecast horizons. A problem that arises in each of these studies is that it is not *a priori* clear which variables should be included in the forecasting model. While hundreds of macro variables are readily available, computational feasibility traditionally limits the analysis to a small set of selected or pre-screened variables.

The latter approach to address the degrees of freedom problem relies on dimensionality reduction methods that aggregate the data space. Dynamic factor models retrieve a common autoregressive trend from a small number of selected explanatory variables (Stock and Watson, 1989, 1992); principal component analysis extracts multiple static factors from a large set of predictors.[1] Fossati (2015) uses dynamic factors estimated from small panels of macro data, Chen *et al.* (2011) and Fossati (2015) employ factors estimated from large sets of macro data to predict US recessions. Gaudreault *et al.* (2003) and Fossati *et al.* (2017) use dynamic factors retrieved from a small set of selected variables to predict recessions in Canada. While large data macro factors allow for a non-discriminatory approach to variable selection, extracted factors may have little correlation with the targeted recession variable. If the macro factor is extracted from a small set of targeted predictors, the researcher again faces the problem of which variables to include in the estimation.

Recently, two methods have been introduced to the discipline of econometrics that address the conundrum of variable selection by performing estimation and variable selection simultaneously: gradient boosting and penalized regression.[2] Penalized regression adds a penalty term to the estimation's likelihood function that lets coefficients of irrelevant variables shrink to zero. The method applied in this paper is the logistic lasso (least absolute shrinkage operator) developed by Tibshirani (1996). Applications of the lasso in the economic forecasting literature are still rare. Notable exemptions are Li and Chen (2014) who use lasso models to forecast several macroeconomic variables and Bai and Ng (2008) who use lasso as a soft thresholding rule to pre-screen a data set to be used in the estimation of macro factors.

While lasso selects relevant variables within a model, boosting obtains its prediction by averaging over many distinct models. Boosting "combines models that do not perform particularly well individually into one with much improved properties" (Ng, 2014). Recently, Ng (2014) and Berge (2015) have used boosting to forecast US

[1] Stock and Watson (2016) give an overview.
[2] See Varian (2014) for an introduction.

recessions.

In this paper, I employ several specifications of the lasso and boosting models to forecast Canadian recessions at different forecast horizons. Each specification is evaluated with respect to its predictive accuracy in an out-of-sample forecasting exercise. A non-discriminatory approach to data selection allows me to identify new variables that serve as important leading indicators and have previously been ignored in the literature. I further demonstrate how the optimal forecasting model changes over time, providing evidence of structural instability. The paper should also be seen as an exploration into forecasting economic variables in a big data environment, as the methods discussed here can easily be applied to related problems.

The rest of this paper is organized as follows. Section 2.2 discusses the data. Section 2.3 introduces lasso and boosting. The main difference between these models and those conventionally used in econometrics is that lasso and boosting utilize supervised learning, i.e. a model's performance feeds back into estimation of the model's parameters. Section 2.4 describes several versions of this data driven approach to model selection in detail. Section 2.5 examines the models' forecasting abilities. In a pseudo real-time experiment, historical recession probabilities are estimated utilizing only data that would have been available at the point the forecast is made. The section also includes extensions to the lasso and boosting models that further improve the fit. Section 2.6 examines the selected variables of the default lasso and boosting model. Section 2.7 concludes the paper.

2.2 Data

I employ a large set of macroeconomic data including 134 monthly variables for the US economy and 445 monthly variables for the Canadian economy.[3] The variables are transformed to ensure stationarity and standardized as described in the appendix

[3]The inclusion of US data is motivated by the notion of spillovers from the US economy (Beaton *et al.*, 2014). More specifically, Bernard and Gerlach (1998) find that the US yield spread adds significant predictive power to forecasts of Canadian recessions at medium and long forecast horizons.

(section A.2). The final data set includes twelve lags of each variable adjusted for publication lag. This means that if a variable is not publicly available for two month after its realization, the variable itself and the first lag of the variable are omitted from the data set, but lag two to twelve are included. The data is partitioned into a training set that is used to initially estimate the models, and a test or hold-out set that is used to recursively produce out-of-sample forecasts. The training set starts in 1968:2 and includes 142 observations up to 1979:12. The hold-out set starts in 1980:1 and includes 420 observations up to 2014:12.

Combining the Canadian and US data sets yields a $x_{T \times K}$ matrix of $K = 5764$ potential explanatory variables and $T = 562$ time periods. While publication lag is taken into account, data revisions are not. Unfortunately, different vintages of the data are not available for the Canadian economy. This is no doubt a shortcoming, however Chauvet and Piger (2008) find that there is no significant advantage of using real time data over revised data when identifying recessions. Next, I will briefly discuss the Canadian data, the US data and the reference series.

All Canadian data are retrieved from Statstics Canada's CanSim database. As the models are supposed to select relevant variables themselves, my approach to data selection is non-discriminatory. I collect all economic series that are available monthly over the time frame and are still being updated. I limit my analysis to economic variables at the federal level. A multitude of series are available at the provincial level. However, this data is not included in the analysis.[4] The series are then tested for duplicates and ordered into the same groups as the US data. I construct one additional group for Canada featuring prices and production in the energy sector: 'Oil & Gas'. As is standard in the forecasting literature, I additionally construct four yield spreads as the difference between government bond yields at different maturities and the three month prime corporate paper rate. To obtain a continuous gdp variable,

[4]While a provincial analysis could be interesting and a gateway to future research, I assume that significant changes at the provincial level are reflected in changes of variables at the federal level.

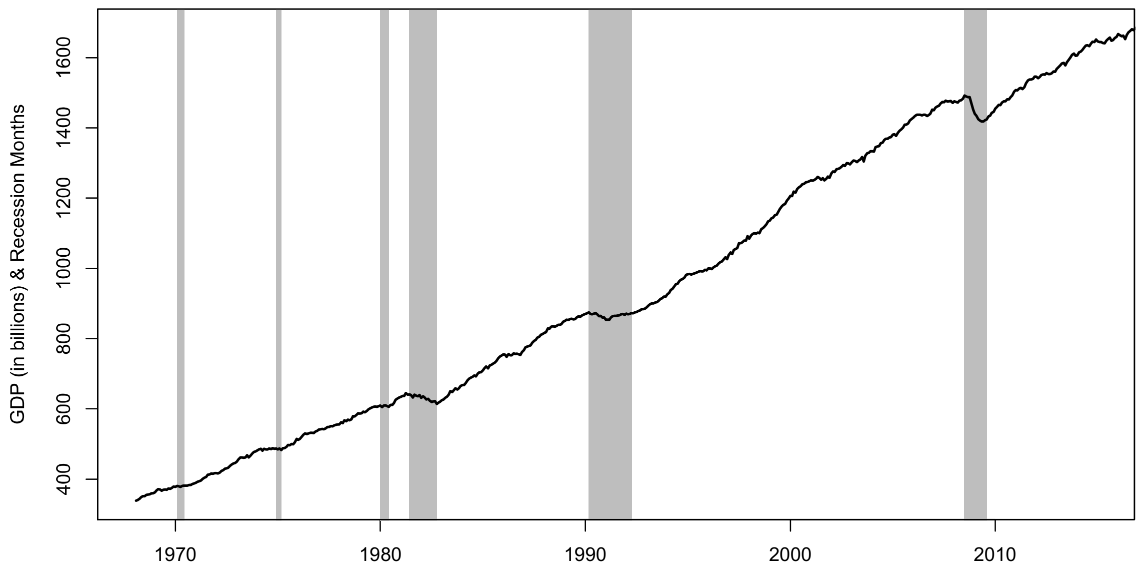

Figure 2.1: Canadian GDP and C.D. Howe recession months illustrated as shaded areas

two different series (v329529 and v65201483) need to be spliced together.[5]

US data is retrieved from the FRED-M monthly database hosted by the Federal Reserve Bank of St. Louis. The data set is regularly updated and publicly available. McCracken and Ng (2014) describe the variables and do some preliminary analysis. The data set including 135 macro series is widely used in macro-econometrics.[6] All data is sorted into the following eight groups: (1) Output and Income, (2) Labour Market, (3) Consumption and Orders, (4) Orders and Sales, (5) Money & Credit, (6) Interest and Exchange Rates, (7) Prices, (8) Stock Market.

Recession months in Canada are determined by the C.D. Howe institute's business cycle council (Business Cycle Council, 2012). y_t is the binary reference series taking on values of 1 for a recession month and 0 otherwise. The C.D. Howe institute publishes recession assessments with considerable delays. I assume 2014:12 to be the last known non-recession month after the last recession in 2008/2009. Figure 2.1 depicts recession months as shaded areas. The figure shows how recessions are closely associated with sharp declines in GDP.

[5]The complete data set is available on my website: http://bit.do/maxsties.
[6]See De Nicolò and Lucchetta (2016) for a recent example.

Cross and Bergevin (2012) discuss each recession and the methodology employed by the Business Cycle Council in detail. They rely on a mix of output and employment variables as well as subjective assessment of historical events such as the bankruptcy of Lehman Brothers in September 2008.

2.3 Two Supervised Learning Methods

This section introduces the lasso and boosting algorithms. Lasso starts out with the complete set of $K = 5764$ variables eliminating non-relevant variables by means of penalization. On the contrary, boosting successively adds variables that add explanatory power. Each method is first described analytically followed by an illustrative example using the full sample for estimation. The mechanics described in this section and the following section 2.4 are working in the background at each stage of the recursive out-of-sample estimation discussed in section 2.5.

2.3.1 Lasso

The least absolute shrinkage and selection operator (lasso) was developed by Tibshirani (1996). It adds a penalty term to the likelihood function of regressions. For binary applications like forecasting recessions, Lokhorst (1999) and Shevade and Keerthi (2003) develop a lasso for logistic regression. The logistic regression model derives the probability of $y_t = 1$ as

$$p(y_t = 1|x_t) = \hat{p}_t = \frac{e^{(\beta_0 + x'\beta)}}{1 + e^{(\beta_0 + x'\beta)}}$$

With a $T \times K$ matrix of time series covariates x, regressing a $T \times 1$ reference series y on x via penalized logistic regression can be represented by minimizing the penalized negative likelihood

$$\min_{\beta_0,\beta} \left\{ -\sum_{t=1}^{T} \left(y_t(\beta_0 + x_t'\beta) - \ln(1 + e^{(\beta_0 + x_t'\beta)}) \right) + \lambda \sum_{k=1}^{K} |\beta_k| \right\},$$

where λ is called the regularization parameter that controls the amount of penalization. $\sum_{k=1}^{K} |\beta_k| = P$ is the non-concave $L1$ norm penalty term on the coefficients.

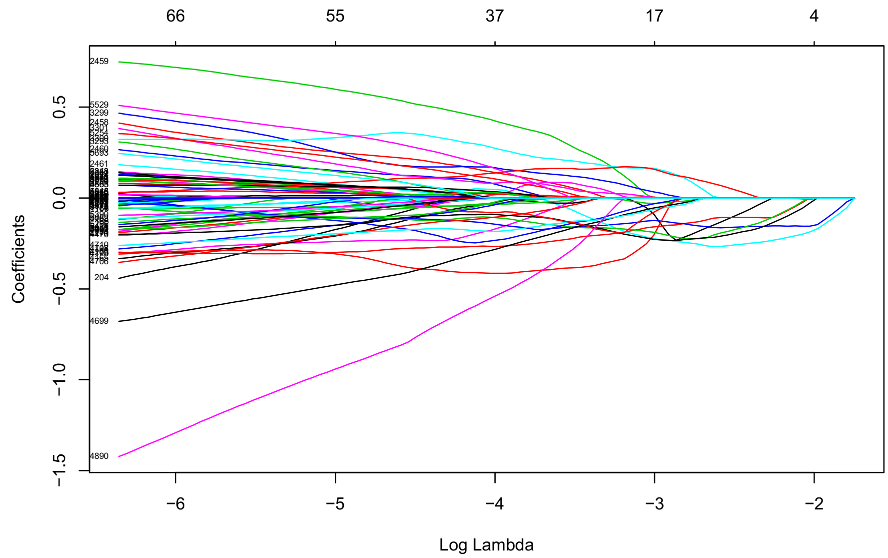

Figure 2.2: Coefficients at different logarithmized values of the regularization parameter λ derived from full sample lasso estimation. The upper x-axis depicts associated degrees of freedom, i.e. number of variables with non-zero coefficient. Each line illustrates the coefficients of a certain variable.

Solving for β_k is likely to lead to corner solutions making the respective coefficient zero. Only coefficients that contribute significantly to the likelihood function will be non-zero. The set of non-zero coefficients therefore determines the shrunk set of *relevant* predictors. Note that the regularization parameter λ needs to be chosen by the researcher. If λ is chosen to be equal to zero, the penalization term drops out of the equation and we are left with regular logistic regression including a non-zero coefficient on each of the K variables.

Section 2.5.2 also presents the results of two modifications of the default lasso model that each feature an alternate penalty term. (1) The elastic net estimator lets the penalty term vary between an L1 and an L2 norm, such that $P_\alpha = \sum_{k=1}^{K} \left(\frac{1}{2}(1-\alpha)\beta_k^2 + \alpha|\beta_k|\right)$ (Zou and Hastie, 2005). The relative weight given to the L1 penalty term, α, can be chosen by validation or set arbitrarily between 0 and 1. (2) The adaptive lasso puts a different weight ω_k on the penalty of each coefficient during estimation such that $P_{ada} = \sum_{k=1}^{K} \hat{\omega}_k|\beta_k|$, where ω needs to be the a monotone transformation of the inverse of a consistent estimator of β, say $\hat{\beta}$ (Zou, 2006). In particular $\hat{\omega} = \frac{1}{|\hat{\beta}|^{\gamma}}$, where γ is a positive constant. This modification makes the coefficients model selection consistent meaning that as $T \longrightarrow \infty$, the procedure consistently picks the correct number of relevant predictors as long as the correlation between relevant and irrelevant predictors is not too large.

Lasso regression is implemented using the *glmnet* package in R that approximates the solution to this maximization problem using coordinate decent (Friedman *et al.*, 2010).

As an illustrating example, figure 2.2 plots the magnitude of the lasso coefficients, estimated from the full sample, against a sequence of the logarithmized regularization parameter λ. On the upper axis, the graph also displays the associated number of non-zero coefficients. Models with larger λ will have a heavier penalty on the size of the coefficients leading to fewer non-zero coefficients. As λ approaches zero, all coefficients become non-zero and the lasso estimator approaches the maximum likelihood estimator.

The best model will balance the out-of-sample error due to bias of robust but small models against the out-of-sample error due to variance of large but overfit models. This model is usually found by evaluating each model's quasi-out-of-sample performance. Section 2.4 discusses such validation procedures in detail.

2.3.2 Boosting

Boosting is a decision tree based method developed by Schapire (1990) and Freund *et al.* (1996).[7] Unlike other tree based methods where decision trees are usually independent, boosting sequentially adds the tree with the best conditional predictive power. The idea is to give observations that were falsely predicted in previous estimation steps more weight in the current step. This boils down to sequentially finding regression trees that best fit the residual variation in the data not explained

[7]Friedman *et al.* (2000) and Mayr *et al.* (2014) provide overviews of the wide array of boosting algorithms and their applications.

in previous steps.

But before the boosting algorithm is discussed in further detail, it is useful to understand the method of regression trees upon which boosting is built. Instead of linking an input variable x^j with a binary reference series y linearly via coefficients, regression trees simply split a given variable x^j into two regions, where each of the regions either produces a positive signal ($\hat{p}_t = 1$) or negative signal ($\hat{p}_t = 0$). At each of these decision nodes, the method simultaneously determines the optimal splitting variable x^j and the optimal splitting point c^j. Regression trees are usually made up of several such decision nodes with each node determining a cut-off for either a new variable or and additional cut-off for a variable already in the tree. For the default boosting model only single node, and therefore univariate, regression trees are considered. In this case, decision trees become decision "stumps" and can be represented as a simple indicator function

$$g(x_t^j) = I(x^j, c^j).$$

Depending on the sign of the correlation of x^j with the reference series, the function produces a positive or negative signal $\hat{p}_t$ when observation x_t^j falls below or above the threshold c^j. Boosting simply adds many of these regression trees together.

In particular, logistic gradient boosting initializes the predictive function $\hat{f}(x)$ as the constant log odds ratio,

$$\hat{f}^0(x) = \log \frac{\sum_{t=1}^{T} y_t}{\sum_{t=1}^{T} (1 - y_t)}.$$

For each of steps $m = 1 \ldots M$, the algorithm then derives the residual variation not explained by the previous model $\hat{f}^{m-1}(x_t)$ as

$$z_t^m = y_t - \frac{1}{1 + e^{\hat{f}^{m-1}(x_t)}} = y_t - \hat{p}_t^{m-1},$$

where $\hat{p}_t^{m-1}$ are the predicted values from model $\hat{f}^{m-1}(x_t)$. The algorithm then fits a univariate regression tree with a single split on the residuals z_t^m,

$$g^m(x_t^j) = E(z_t^m | x_t).$$

Table 2.1: Demonstration of Boosting Algorithm

	SplitVar x^j	SplitCodePred c^j	ErrorReduction
Step 1:	$itre3_3$	0.16	19.35
Step 2:	$REALLN_{12}$	-0.86	6.44
Step 3:	$CLAIMSx_{12}$	-2.40	4.29
Step 4:	$fpus3_{11}$	2.62	3.35
Step 5:	$NONREVSL_5$	-1.75	2.07
Step 6:	$IPFUELS_{12}$	-1.90	2.26
Step 7:	$itre3_6$	0.21	1.49
Step 8:	$NONREVSL_3$	-1.13	1.51
Step 9:	$exswe_7$	2.87	1.11
Step 10:	$itre3_1$	0.17	1.27
Step 11:	$CES3000000008_{12}$	-2.29	0.88
Step 12:	$NONREVSL_6$	-1.45	0.94
Step 13:	$USWTRADE_{12}$	-2.37	0.65
Step 14:	$itre3_1$	0.16	0.62
Step 15:	$NONREVSL_4$	-1.13	0.55
Step 16:	$houstot_{12}$	-1.15	0.45
Step 17:	$exswe_7$	2.87	0.36
Step 18:	$bankacp1m_{12}$	1.07	0.31
Step 19:	$USGOOD_1$	-2.90	0.29
Step 20:	$itre3_1$	0.18	0.28

Notes: Demonstration of boosting algorithm using $M = 20$ steps, a learning rate $\mu = 0.5$ and forecast horizon $h = 1$. 'SplitVar' refers to the optimal splitting variable in the respective step x^j, 'SplitCodePred' indicates the optimal cut-off for the variable c^j, and 'ErrorReduction' refers to the decline in the negative of the binomial likelihood. Variable names using capital letters refer to US variables, subscripts refer to variable lag. A short description of the variables can be found in appendix A.1.

As discussed above, $g(x_j)$ partitions the predictor space into disjoint regions. At each step, the model selects the variable that best fits the residuals from the previous step and estimates a cut-off that determines a positive (recession) or negative (no recession) signal for each observation. Using univariate regression trees at each boosting step is called component-wise boosting (Bühlmann and Yu, 2003; Ng, 2014) and allows for better interpretability of boosting as a variable selection tool. The optimal variable x_t^j as well as the optimal cut-off c^j are found simultaneously by minimizing a loss function Ψ,

$$\hat{g}^m(x_t^j) = \arg\min \sum_{t=1}^{T} \Psi\left(y_t, \hat{f}^{m-1}(x_t) + \hat{g}^m(x_t^j)\right).$$

The particular loss function minimized in this application is the negative binominal log likelihood. Finally, the regression tree found in step m is added to the model from step $m-1$ scaled down by the learning rate parameter μ,

$$\hat{f}^m(x) = \hat{f}^{m-1}(x) + \mu\hat{g}^m(x).$$

By combining M individual regression trees this way, multiple weak learners, individual regression trees, build a strong learner, the function $\hat{f}^M(x)$. The learning rate specifies how big the steps are that taken while building the aggregate model. A low learning rate is crucial to avoid overfitting. The complete boosting algorithm employed in this paper can be found in appendix A.3.

Section 2.5.2 also presents two modifications of the default boosting model. (1) Friedman (2002) introduces stochastic boosting where new trees are fit on a subsample of the data. This is called bagging and improves the robustness of the model. (2) Pairwise boosting increases the interaction depth to two, such that at each stage of the boosting algorithm the best two node regression tree is fit on the residuals of the previous step.

Boosting is implemented in R using the *gbm* package. I specify the maximum number of steps as $M^{max} = 1000$ and the learning rate as $\mu = 0.01$ which are standard in the literature.

Table 2.1 illustrates an example of the boosting algorithm using the full sample for $h = 1$ month ahead forecasts. In the first step, the third lag of the Canadian 3-month treasury bill yield, $itre3_3$, is selected as best predictor. If the standardized value of the treasury bill yield exceeds the cut-off value of $c^j = 0.16$, the model predicts a positive recession signal scaled down by the learning rate $\mu = 0.5$, meaning that only half of the predicting generated in step m is added to the aggregated mode. The initial regression tree $g^1(x_j)$ reduces the loss function, the negative of the binomial likelihood, by 19.35 points. The 12th lag of US real estate loans at commercial banks, $REALLN_{12}$, best fits the residuals not explained in the first step. When the variable falls below a cut-off of $c^j = -0.86$, an additional recession signal is added. Note that some variables appear repeatedly throughout the 20 steps. Each additional regression tree is able to reduce the binomial likelihood, but at an increasingly smaller amount. In the last step, the binomial likelihood is only reduced by an additional 0.28 points.

As the case with lasso, boosting is prone to overfitting. One measure to counter overfitting is setting the learning rate μ low to slow down the algorithm. However, the most important variable and the equivalent of λ for lasso, is the number of trees or models fit sequentially, M. With a small number of steps M one obtains a highly biased, low variance model with only few predictors. If M is large the model fits the data almost perfectly in sample, however the model carries high variance and is useless for out-of-sample forecasting. As with the regularization parameter λ for lasso, the optimal number of iterations M is chosen by validation, as discussed in the next section.

2.4 Model Validation

In this section, I discuss several validation procedures in the context of recession forecasting. A detailed survey on these and other cross validation methods is provided by Arlot *et al.* (2010). The role of validation is to determine which among the sequence of models featuring distinct number of variables should be used for out-of-sample forecasting. Each procedure estimates the model for a range of the regularization parameter (λ for lasso, M for boosting) and selects the model with the best quasi-out-of-sample performance. This means that for each value of the regularization parameter the model is estimated on a subset of observations to predict the left out observations of the reference series. The optimal value of the regularization parameter is the one that is associated with the model that predicts the left out observations with the smallest amount of error. This prediction error can be calculated using various goodness of fit measures that are introduced next.

2.4.1 Goodness of Fit Measures

This subsection introduces potential goodness of fit measures for probability predictions of a binary reference series. For each measure, lower values indicate better fit.[8]

QPS. The quadratic probability score (QPS) is defined as

$$\text{QPS} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{p}_t)^2$$

and is the most standard measure. It is simply the mean squared error for binary regression. The QPS can take on values between zero and one.

LPS. The logarithmic probability score (LPS) is defined as

$$\text{LPS} = -\frac{2}{T}\sum_{t=1}^{T}\left(y_t \log(\hat{p}_t) + (1 - y_t)\log(1 - \hat{p}_t)\right).$$

This is -2 times the log likelihood or log deviation. The LPS can also be seen as a variation of the QPS that puts a larger weight on large deviations between estimated probabilities and the reference series. It can take on values between zero and infinity.

MCE. The misclassification error is the share of false predictions. This includes false positives as well was false negatives.

$$\text{MCE} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(y_t \neq f(\hat{p}_t))$$

[8]For better comparability, the AUC measure is interpreted as 1-‘area under the ROC curve’ as larger values of the area under the curve indicate better fit. See appendix A.4 for details.

where

$$f(\hat{p}_t) = \begin{cases} 1 & \text{if } \hat{p}_t \geq c \\ 0 & \text{otherwise} \end{cases}$$

To calculate the misclassification error in practice, one needs to determine the threshold c that decides at which point a probability is counted as a positive signal. Two natural choices arise. A value of $c = 0.5$ means that if the estimated probability of recession exceeds 1/2, a recession is indicated (Chauvet and Potter, 2010). Another choice for c is the sample mean $c = \frac{\sum_{t=1}^{T} y_t}{T}$ used in this paper. Cramer (1999) argues that the sample mean is the appropriate cut-off for unbalanced samples like the one employed in this paper.[9] However, it should be noted that the choice of c is arbitrary.

AUC. The area under the receiver-operator curve or ROC takes account of this arbitration. It is defined as

$$\text{AUC} = 1 - \int_{\infty}^{-\infty} \text{TPR}(c)\text{FPR}'(c),$$

where TPR is the true positive rate and FPR is the false positive rate. The AUC generalizes the classification error over all possible thresholds c and has previously been employed by Berge and Jorda (2011) in the context of recession forecasts.[10]

2.4.2 k-fold Cross Validation

The most common validation procedure in the machine learning literature is k-fold cross validation. This entails dividing the sample into k (possibly random) subsamples, using $k-1$ subsamples to fit the model and produce predictions for the k^{th} subsample. Once predictions are obtained for the entire sample using this quasi-out-of-sample exercise repeatedly, the joint set of predictions can be compared to the reference series using a goodness of fit measure. Let the predictions obtained from k-fold cross validation be denoted by $\hat{p}_t^{cv}$, then the cross validation error CV is the

[9] 'Unbalanced sample' in this context refers to an unequal distribution of positive and negative events in the reference series.

[10] More on the AUC can be found in appendix A.4.

average of the goodness of fit measure over all k subsamples,

$$CV^{GF} = \sum_{\kappa=1}^{k} \frac{1}{k} GF_{\kappa}.$$

Repeating this procedure for a sequence of values of the regularization parameter will give a sequence of associated cross validation errors. The minimum cross validation error in this sequence indicates the optimal value of the regularization parameter that produces the lowest out-of-sample error. It is common to choose the highest value of

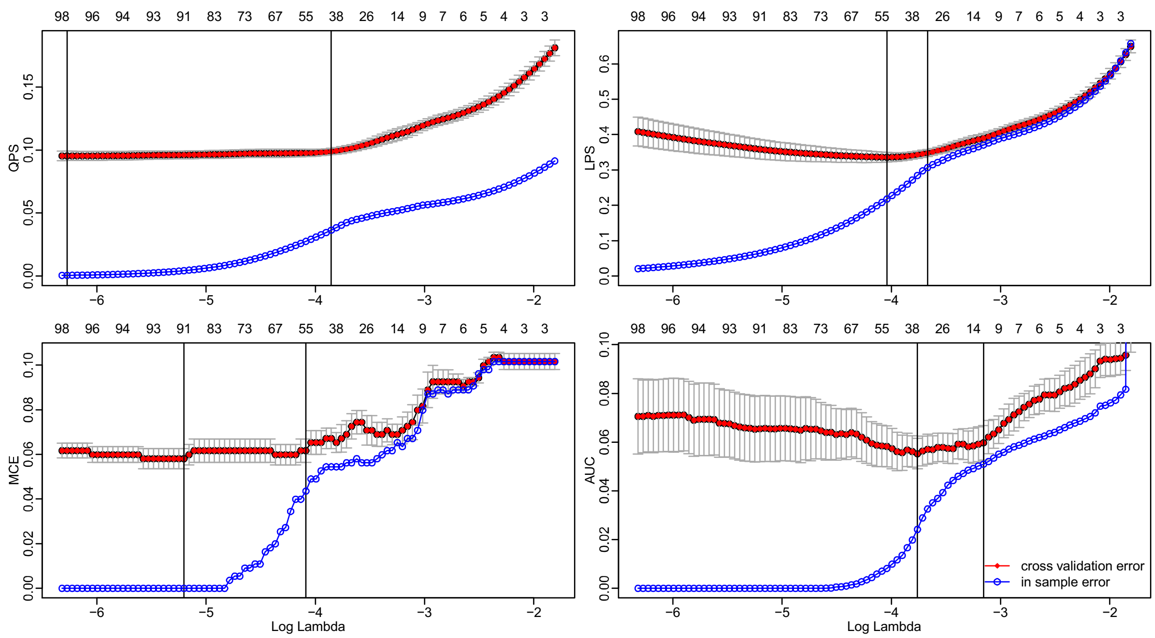

Figure 2.3: Cross validation error using different goodness of fit measures derived from full sample lasso estimation at different values for the logarithmized regularization parameter λ. The upper x-axis depicts the associated number of variables with non-zero coefficient. The first dotted line from the left indicates the minimum CV, the second dotted line indicates the smallest CV within one standard error of the minimum CV. The blue line plots the corresponding in-sample error.

λ within one standard deviation of the cross validation error to account for the fact that cross validation predictions are obtained from estimations with lower sample size than the full model and err on the side of more parsimonious models. Standard errors can be obtained as

$$SE = \frac{SD_k}{\sqrt{k}}.$$

To illustrate this procedure, figure 2.3 plots the 12-step ahead CV of full sample lasso estimation using several goodness of fit measures against a sequence of lambda. The (first) global minimum of each cross validation curve indicates the optimal level of lambda. Different goodness of fit or error measures indicate different optimal values for λ as discussed in the next subsection. Note also that the in-sample error keeps decreasing from right to left and eventually approaches zero when lambda is chosen small enough. Once enough variables are added, the resulting model can fit the reference series perfectly in sample.

2.4.3 k-step Ahead Time Series Cross Validation.

For the prediction of macroeconomic time-series, k-fold cross validation can be modified to better reflect the nature of time series data. As with regular cross validation, the sample is divided into several subsamples, however, to obtain the time series cross validation error ($TSCV$), the sample is divided in a $1, \ldots, t_{ini} = 90$ observation initialization sample and a validation sample $(t_{ini} + 1), \ldots, T = 562$. The validation sample is then divided into $\frac{T-t_{ini}}{k} = S$ continuous subsamples. Note that k now denotes the number of observations in a subsample and not the number of subsamples. A prediction is then made for each period in the validation sample using a model estimated from the initialization sample and all subsamples that occur before the period for which the prediction is made. The $TSCV$ can be obtained as[11]

$$TSCV^{GF} = \sum_{s=1}^{S} \frac{k}{sk} GF_k.$$

The 'within one standard error rule' should not be applied when using time series cross validation for two reasons. First, the TSCV already reflects the real out-of-sample error much closer and chooses more robust models with fewer variables than regular cross validation. Second, estimated standard errors of the TSCV are generally

[11] Note that if $\frac{T-t_{ini}}{k}$ does not divide by k, some observation at the end of the sample will be left out.

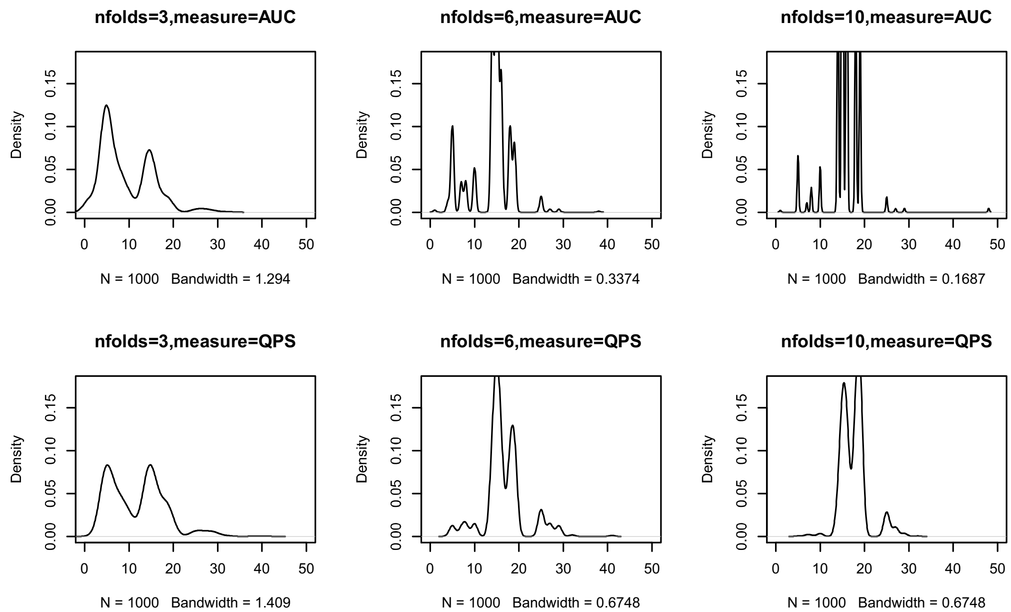

Figure 2.4: Distribution of the optimal number of variables obtained from repeated cross validation of full sample lasso estimation using a different number of validation folds (nfolds) and different goodness of fit measures (measure).

larger, thus picking the smallest model within one standard error can lead to models that only include a minimal number of variables.

2.4.4 Cross Validation with Binary Reference Series

Cross validation with a binary reference series such as a recession indicator brings about several problems. Recessions are rare events and amount for around 12% of the sample period. Additionally, the data suffers from structural instability as different recessions have different origins. When using the standard cross validation procedure, the sample is divided randomly into k subsamples where $k-1$ are used for estimation and the left out subsample is used to evaluate the fit. Due to the structural instability, the models estimated from different random subsets can be quite different depending on which recessions are by chance not included in the estimation sample.

An example of this inconsistency is displayed in figure 2.4. When full sample lasso estimation is repeatedly cross validated using random subsampling with replacement, the optimal number of variables as indicated by the cross validation procedure varies.

Table 2.2: Variability of Cross Validation with Random Subsampling

	AUC3	AUC6	AUC10	DEV3	DEV6	DEV10
Mean	9.1	13.8	15.2	11.6	16.6	17.6
SD	5.7	4.4	2.9	6.4	4.6	3.0

Notes: Mean and standard deviation of the optimal number of variables associated with the cross validation procedures described in figure 2.4.

The figure demonstrates that a wide range of outcomes is possible for different numbers of cross validation folds k as well as different goodness of fit measures. The problem is somewhat mitigated with a larger amount of cross validation folds. I address this problem by determining the split of observations for k-fold cross validation *a priori*. The k subsamples are constructed such that subsequent observations are assigned to different subsets. Specifically, when x_t and y_t are assigned to fold k_i, observation x_{t+1} and y_{t+1} are assigned to fold k_{i+1}. This guarantees the most even split of recession periods between the k cross validation folds and eliminates the randomness.

The second problem is that as recessions only make up 12% of the sample, the selected model will be biased towards predicting non-recession months if the average deviation from the reference series is used as a goodness of fit measure as is the case with the LPS and QPS. This can be counteracted by either giving recession and non-recession months equal weight in the cross validation procedure or by making use of classification type errors such as the MCE or AUC.

Alternatively, time series cross validation can be applied. This also eliminates randomness, however due to structural instability, models estimated from different number of folds change significantly. Additionally, the observations in each validation subsample are consecutive, making it possible for validation subsamples to include only non-recession months. The AUC can therefore not be calculated for time series cross validation as the measure cannot be calculated without positive events in the validation sample. Finally, when the sample size is small and barely exceeds the initialization sample t_{ini}, not enough validation samples exist to form a measure of the standard error. The 'within one standard error' rule can therefore not be applied. Section 2.5.2 compares the out-of-sample forecast using time series cross validation to the results obtained from standard cross validation.

2.5 Estimation Results

This section presents historic recession probabilities obtained by direct recursive out-of-sample forecasting. To determine the out-of-sample forecasting power of the lasso and boosting algorithms, I divide the data into a training and a hold-out sample. For each observation in the hold-out sample a forecast is produced using only data up to the point in time the forecast is made. The estimation using all data including the last observation in the hold-out sample naturally is equivalent to in-sample or full sample estimation. The following subsection presents the results for the default lasso and boosting model, followed by a subsection examining the fit of alternative model specifications.

2.5.1 Recursive Out-of-Sample Forecasting

For each observation in the hold-out sample I produce direct h step ahead forecasts for $h = (1, 3, 6, 12)$ made at the end of each month. A forecast is equivalent to the probability of a certain observation in the hold-out period being a recession month. Assume lasso and boosting can be reduced to produce a predicted probability through the function $F(y, x)$ as described in section 2.3, then the probability of month t being a recession month is determined by

$$Pr(y_{t+h} = 1|x_t) = \hat{p}_{t+h} = F(y_t, x_{t-h}),$$

where x_t includes up to 12 lags of each variable. A 6-step ahead probability forecast $\hat{p}_{t+h}$ for 1980:7 made in t =1980:1, uses data x_{t-h} up to 1979:7 and the reference series y_t up to 1980:1, to estimate the model.

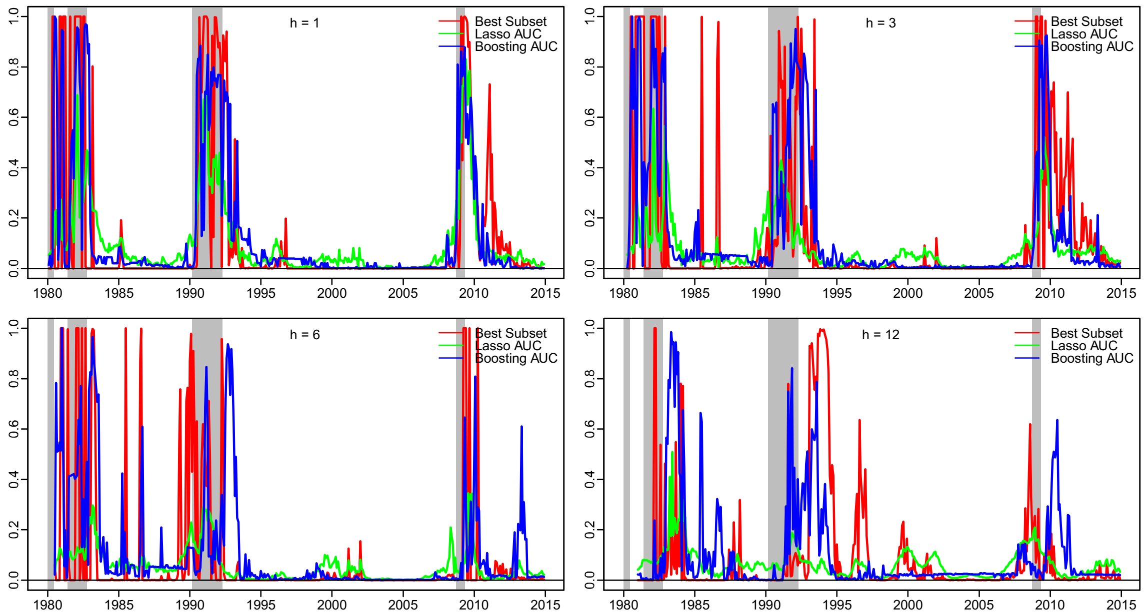

Figure 2.5: Estimated out-of-sample probabilities of the default lasso and boosting models validated by cross validation at different forecast horizons (h). Best Subset depicts out-of-sample probabilities derived from best subset selection logit regression.

Figure 2.5 depicts estimated out-of-sample recession probabilities for the standard lasso and boosting models using 5-fold AUC cross validation. The models are contrasted to a naive best subset selection model that chooses a model of up to five variables from a subset of 15 variables obtained from hard thresholding as described in Bai and Ng (2008).

Boosting and best subset appear to be more responsive than the lasso model that produces lower recession probabilities throughout the sample. On the other hand, the lasso model also produces fewer false positive predictions. Note that especially for boosting, estimated probabilities seem to spike with a delay of h. This is due to the structural instability of the model. Some recessions are only recognized once the according observation becomes part of the estimation since variables that were good predictors in previous recessions are not necessarily good predictors for the current recession. If a recession starts at t, in $t+h$ the first observation of the recession is included in the estimation and the out-of-sample estimates indicate an increased recession probability. This naturally makes for a bad forecast. Subsection 2.5.2

Table 2.3: Goodness of Fit of Default Models

	bss.15.5	boost.cv.5.auc	lasso.cv.5.auc
		-h0-	
QPS	0.02	**0.01**	0.05
LPS	0.15	**0.11**	0.32
MCE	0.06	**0.04**	0.07
AUC	0.01	**0**	0.02
n.var	5	61	44
		-h1-	
QPS	0.09	0.07	**0.07**
LPS	2.76	0.58	**0.46**
MCE	0.14	0.11	**0.10**
AUC	0.20	0.10	**0.05**
n.var	5	59	39
		-h3-	
QPS	0.13	0.09	**0.09**
LPS	4.04	0.69	**0.58**
MCE	0.19	0.13	**0.12**
AUC	0.28	0.14	**0.08**
n.var	4	9	31
		-h6-	
QPS	0.13	0.13	**0.10**
LPS	3.28	1	**0.65**
MCE	0.14	0.19	**0.13**
AUC	0.30	0.25	**0.12**
n.var	5	15	50
		-h12-	
QPS	0.16	0.15	**0.11**
LPS	3.28	1.48	**0.74**
MCE	0.23	0.27	**0.15**
AUC	0.53	0.57	**0.22**
n.var	5	7	22

Notes: 'bss.15.5' refers to best subset selection choosing the best model of up to 5 variables from a set of 15 preselected potential predictors. The default lasso and boosting models are validated via 5-fold cross validation using the auc as goodness of fit measure for validation. 'n.var' refers to the number of non-zero coefficients obtained from full sample estimation.

discusses alternative model specifications that exhibit better fit. Table 2.3 presents the goodness of fit for the default models depicted in figure 2.5. Due to the lowest amount of false positive signals, lasso estimates dominate boosting estimates slightly. The number of selected variables will change over the sample period as the models are reestimated for each t, so the number of variables stated here is the number of variables used in the last out-of-sample estimation which is equivalent to full sample estimation. The change of selected variables over time is discussed in section 2.6.

2.5.2 Alternative Model Specifications

This subsection discusses alternative model specifications to the default models from 2.5.1. First, I will discuss the effect of different validation procedures on the out-of-sample fit. Then, I will introduce some changes to the lasso and boosting procedure that theoretically improve the fit. Lastly, I will present estimates that take into account that the reference series is only available with substantial delay.

Alternative Validation Procedures. As discussed in 2.4, different validation procedures will lead to different numbers of variables being picked within the same model family. Note that within cross validation, the variation in out-of-sample forecasts is minimal. Figure 2.6 shows that the probabilities from 5-fold and 10-fold cross validation as well as using AUC or LPS for validation are nearly identical. The same results are true for Boosting but to a lesser extent.

Time series cross validation does not yield any improvement in fit but does take significantly longer to estimate and leads to the problems discussed in 2.4.4.

Adaptive Lasso and Elastic Net. As described in section 2.3.1, the default lasso is not model selection consistent. Two modification should therefore yield an improvement of out-of-sample fit. The elastic net estimator is an extension of the lasso that allows for a mix between L1 and L2 penalty. α is the proportion of L1 penalty in the estimation. I choose α by cross validation so the optimal alpha varies over the out-of-sample period. I also set α arbitrarily equal to 0.6 and 0.8. The adaptive lasso

Table 2.4: Goodness of Fit Using Various Validation Methods

	l.cv.5.lps	l.cv.10.lps	l.cv.5.auc	l.cv.10.auc	l.tscv.12.lps	l.tscv.60.lps	b.cv.5.lps	b.cv.10.lps	b.cv.5.auc	b.cv.10.auc	b.tscv.12.lps	b.tscv.60.lps
		-h0-										
QPS	0.04	0.04	0.05	0.05	0.05	0.06	0.02	0.04	**0.01**	0.10	0.05	0.06
LPS	0.25	0.30	0.32	0.33	0.34	0.43	0.21	0.32	**0.11**	0.72	0.38	0.42
MCE	0.05	0.05	0.07	0.07	0.07	0.11	0.05	0.14	**0.04**	0.22	0.14	0.15
AUC	0.01	0.02	0.02	0.02	0.01	0.03	0.01	0.03	0	0.22	0.04	0.02
n.var	55	51	44	45	44	44	19	6	61	19	27	14
		-h1-										
QPS	0.07	0.07	0.07	0.07	0.07	0.09	0.07	0.07	0.07	0.11	0.07	0.09
LPS	**0.43**	0.46	0.46	0.46	0.47	0.60	0.50	0.51	0.58	0.70	0.49	0.60
MCE	**0.09**	0.10	0.10	0.10	0.11	0.14	0.12	0.14	0.11	0.19	0.14	0.19
AUC	**0.04**	0.05	0.05	0.05	0.07	0.11	0.11	0.10	0.10	0.17	0.09	0.14
n.var	51	42	39	39	41	41	21	12	59	21	31	17
		-h3-										
QPS	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.10	0.09	0.13	0.10	0.12
LPS	**0.55**	0.57	0.58	0.57	Inf	0.64	0.65	0.66	0.69	0.89	0.63	0.79
MCE	0.12	0.12	0.12	0.13	0.12	0.12	0.16	0.16	0.13	0.24	0.17	0.35
AUC	0.07	0.07	0.08	0.07	0.14	0.13	0.15	0.17	0.14	0.32	0.17	0.37
n.var	45	38	31	47	34	34	9	9	9	9	20	12
		-h6-										
QPS	0.10	0.10	0.10	0.10	0.11	0.11	0.12	0.12	0.13	0.13	0.12	0.11
LPS	0.64	0.65	0.65	0.64		0.82	0.79	0.79	1	0.93	0.80	0.78
MCE	0.13	0.13	0.13	0.14	0.14	0.18	0.18	0.21	0.19	0.25	0.21	0.23
AUC	0.12	0.12	0.12	**0.11**	0.19	0.20	0.22	0.29	0.25	0.36	0.40	0.51
n.var	43	39	50	29	17	17	19	3	15	19	16	12
		-h12-										
QPS	0.11	0.11	0.11	0.11	0.11	0.12	0.15	0.15	0.15	0.15	0.13	0.12
LPS	0.74	**0.73**	0.74	0.74	0.77	0.84	0.95	0.94	1.48	0.99	0.87	0.86
MCE	0.16	0.16	**0.15**	0.16	0.34	0.55	0.29	0.27	0.27	0.28	0.31	0.42
AUC	0.20	0.20	0.22	0.22	0.42	0.80	0.38	0.35	0.57	0.38	0.51	0.86
n.var	44	33	22	44	9	9	7	11	7	7	22	3

Notes: 'l.*' models indicate lasso, 'b.*' indicate boosting models. 'cv' and 'tscv' refer to cross validation and time series cross validation respectively. The number in the model name indicate the number of folds in cross validation or the number of k-step ahead forecasts for time series cross validation. 'lps' and 'auc' indicate the goodness of fit criterion used for validation.

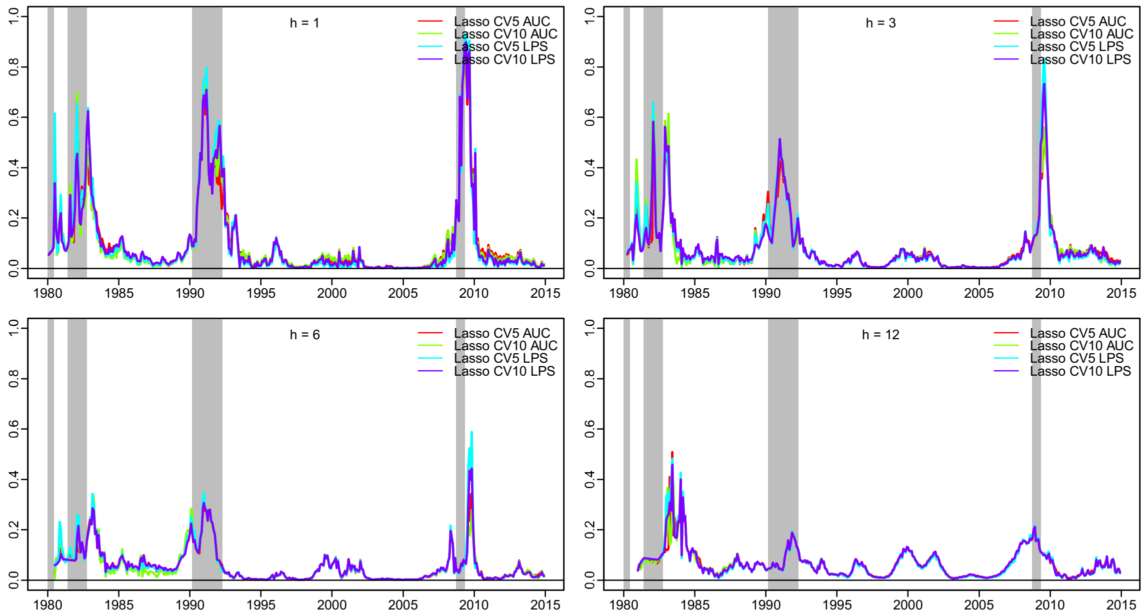

Figure 2.6: Estimated out-of-sample probabilities using 5-fold and 10-fold cross validation with AUC and LPS as selection criterion at different forecast horizons

weighs the penalty term for each coefficient proportional to a consistent estimate of the coefficients. γ determines the strength of the penalty relative to the absolute size of the consistent coefficient estimate. I choose $\gamma \ \epsilon \ [0.5, 0.75, 1]$.

Note in figure 2.7 that there is barely any difference between the default lasso and the elastic net with cross validated α. However, the adaptive lasso is significantly more responsive and the problem of a delayed response to recessions is mitigated. It is noteworthy that all adaptive lasso estimates predict a recession in 1986, when Canada was hit harshly by a decline in oil prices. The C.D. Howe business cycle committee however, does not classify this period as recession. It is known that the adaptive lasso has a relatively high false positive rate. Sampson *et al.* (2013) discuss and address this problem. These false positives lead to higher prediction errors for the adaptive lasso forecasts. For the nowcasts ($h = 0$), the adaptive lasso models however produce near perfect predictions.

Stochastic and Pairwise Boosting. The boosting algorithm can be extended to allow for pairwise interaction between the covariates. The algorithm lets each

Table 2.5: Goodness of Fit Using Alternative Model Specifications

	enet.cv.5.auc	enet.cv.5.lps	enet.alpha.6.auc	enet.alpha.8.auc	lasso.ada.gam.5	lasso.ada.gam.75	lasso.ada.gam1	boost int.d.=2	boost bf=.8
		-h0-							
QPS	0.05	0.04	0.02	0.02	**0**	0.01	0.03	0.02	0.03
LPS	0.32	0.25	0.14	0.13	**0.04**	0.09	0.27	0.16	0.26
MCE	0.06	0.05	0.04	0.03	**0.01**	0.05	0.14	0.04	0.10
AUC	0.02	0.01	0	0	0	0	0.02	0.01	0.01
n.var	191	201	144	115	61	46	40	93	97
		-h1-							
QPS	0.07	0.07	0.06	0.06	0.07	0.08	0.10	0.05	0.06
LPS	0.47	0.43						0.41	0.42
MCE	0.10	0.09	0.08	**0.07**	0.11	0.13	0.22	0.09	0.14
AUC	0.05	0.04	0.04	0.04	0.07	0.09	0.17	0.07	0.08
n.var	240	301	79	65	70	55	43	71	85
		-h3-							
QPS	0.09	0.09	0.09	0.10	0.11	0.12	0.14	0.09	0.09
LPS	0.57	0.55						0.57	0.57
MCE	0.12	0.12	**0.10**	0.11	0.16	0.18	0.30	0.14	0.15
AUC	0.08	**0.07**	0.14	0.14	0.14	0.19	0.28	0.12	0.12
n.var	57	126	110	93	76	72	68	51	79
		-h6-							
QPS	0.11	**0.10**	0.11	0.11	0.13	0.13	0.15	0.11	0.11
LPS	0.68	0.64						0.70	0.67
MCE	0.14	0.13	0.13	0.13	0.17	0.19	0.32	0.17	0.18
AUC	0.15	**0.12**	0.20	0.21	0.21	0.22	0.27	0.17	0.20
n.var	85	126	96	79	90	74	75	60	74
		-h12-							
QPS	0.11	0.11	0.12	0.12	0.19	0.19	0.16	0.12	0.12
LPS	0.76	**0.74**	0.84	0.87				0.82	0.84
MCE	0.18	**0.16**	0.22	0.20	0.28	0.29	0.35	0.23	0.26
AUC	0.26	**0.20**	0.31	0.34	0.42	0.43	0.37	0.34	0.38
n.var	80	87	53	39	51	43	43	42	62

Notes: The first two elastic net models use two-dimensional cross validation to determine α using different goodness of fit measures for validation. The third and fourth model use a predetermined value of α equal to 0.6 and 0.8 respectively. The adaptive lasso models use values of γ equal to 0.5, 0.75 and 1. The last two models refer to boosting with pairwise interaction (interaction depth of two) and boosting with bagging fraction of 0.8, respectively.

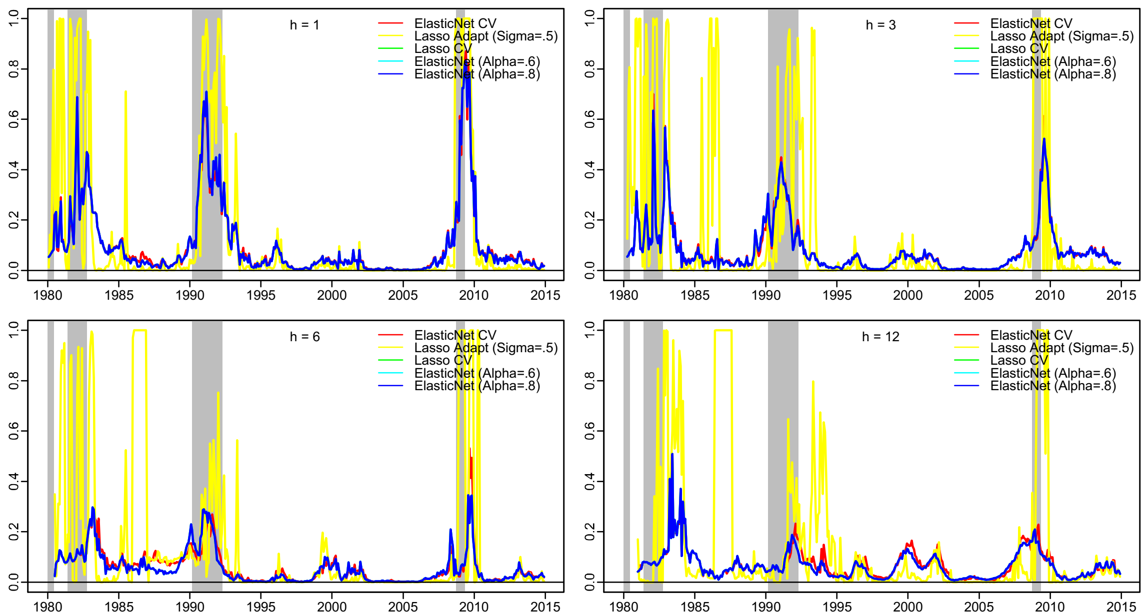

Figure 2.7: Estimated out-of-sample probabilities of the Elastic Net and Adaptive Lasso compared to the default lasso model at different forecast horizons

function derived from a boosting step become a two variable logit regression. This reduces interpretability but potentially increases the model's predictive power. With stochastic boosting, only a fraction of the observations are used for estimation. This is useful to avoid overfitting. Changing the bagging fraction barely alters estimated probabilities. Allowing for pairwise interaction improves the fit slightly but takes significantly longer to estimate.

Publication Lag for the Reference Series. So far the analysis has assumed that up to the current month it is known wether the economy is in a state of recession or not. In reality however, it can take up to one year until recessions are officially declared. With a publication lag of $l = 12$ for y_t, the forecast equation now becomes

$$Pr(y_{t+h} = 1|x_t) = \hat{p}_{t+h} = F(y_{t-l}, x_{t-h-l}).$$

For a forecast horizon of 12 this means that a forecast made at the end of $t = 1980:1$ uses data x_{t-h-l} up to 1978:1 and the reference series y_{t-l} up to 1979:1 to estimate a model that produces a probability forecast $\hat{p}_{t+h}$ for 1981:1. Figure 2.9 depicts the predicted probabilities for the standard models as well as the adaptive lasso. The

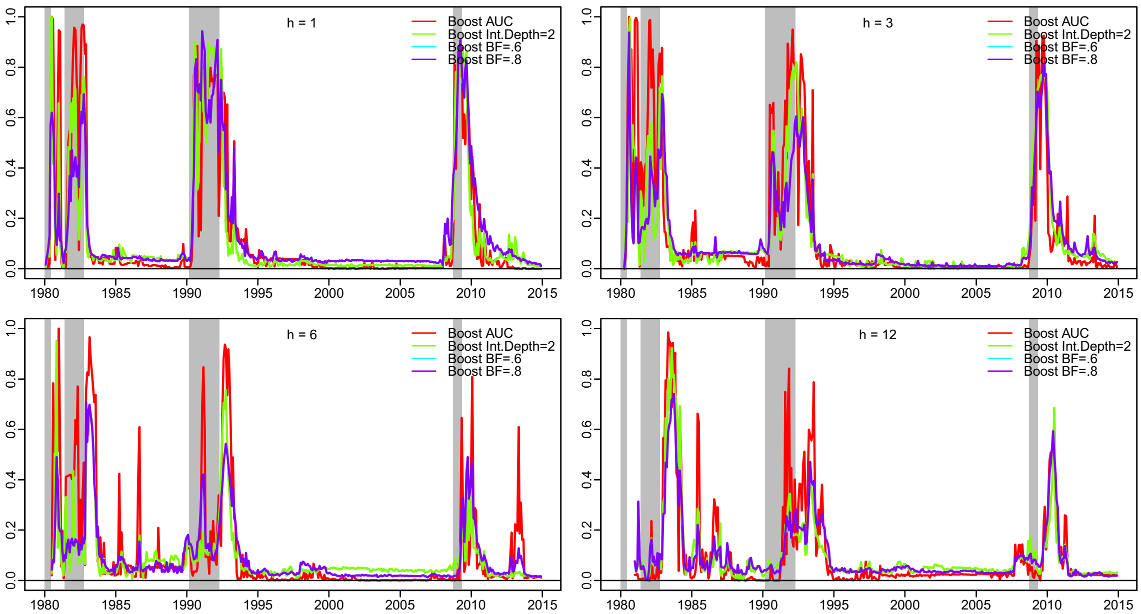

Figure 2.8: Estimated out-of-sample probabilities of stochastic boosting (with bagging fractions (BF) of 0.6 and 0.8) and pairwise boosting compared to the standard boosting model, validated via 5-fold cross validation using the AUC as model selection criterion at different forecast horizons.

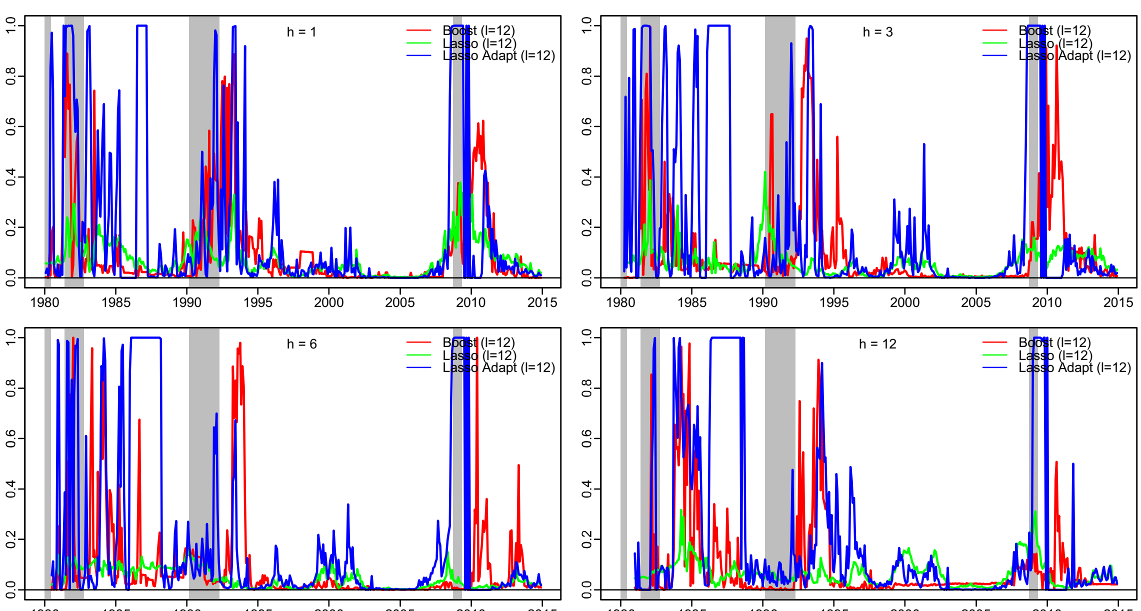

Figure 2.9: Estimated out-of-sample probabilities of default boosting and lasso model, as well as adaptive lasso considering a 12 month publication lag for reference series at different forecast horizons.

nowcast and the 1 month ahead forecast reliably identify recessions with no extended periods of false positives. Predictions at longer forecast horizons give mixed signals with often significantly delayed response and would be hard to interpret in reality.

2.6 Variable Selection

One advantage of not aggregating the data space is in the interpretability of the estimated coefficients. This section presents the variables carrying most conditional predictive power using the default lasso and boosting model. While section 2.6.1 discusses the most important predictors for the complete sample, section 2.6.2 describes the development of this set of variables over time. A short description of the variables presented in this section can be found in the appendix A.1. Before beginning to discuss the selected variables, a word of caution is in order. The analyses in this paper are of purely predictive nature and do not generally allow for causal inference. That being said, the conditional correlation of specific variables with Canadian recessions is worthwhile exploring in itself.

2.6.1 Full Sample Variable Selection

Using all the information available at the end of the sample window makes it possible to identify the most important leading indicators in the data set. While cross validation identifies up to 60 variables with relevant predictive power, most of the data's explanatory power is concentrated in only a handfull of variables. Table 2.6 presents lasso coefficient estimates for the ten variables with largest absolute magnitude. The absolute magnitude of coefficients can be used as a measure of the importance of the variables as all variables have been standardized. For boosting, ranking coefficients is problematic as variables are likely picked in more than one iteration step. Instead, Friedman (2001) constructs a measure of relative importance I_j^2 for variable j. In particular,

$$I_j^2 = \frac{1}{M}\sum_{m=1}^{M} i_m^2 \mathbb{1}(x^m = x_j),$$

where i_m^2 describes the least-squares improvement of x_j in step m. Table 2.7 presents the 10 variables with the highest relative influence in the boosting model. In both tables, capitalized variable names indicate US variables, subscripts indicate lags. In table 2.6, real variables have the highest impact at short horizons. Surprisingly, for the 1-month forecast, the Canadian 10-year yield spread (ys10) is surpassed by US housing market variables (housing starts: HOUST and building permits: PERMITS) in both, the lasso and the boosting model. Unlike lasso, boosting puts the highest weight on 3 month treasury bills (itre3). With employment in transport & utilities, employment in the financial sector, as well as male unemployment (USTPU, USFIRE, USFIRE, UEMP15OV), the US labour market figures prominently into short and medium term forecasts. At longer horizons, yield spreads are the dominant predictors. Additionally to the 5 year yield spread, both forecasting models include IMF currency reserves (resvimf). The boosting model also features the Canadian exchange rate with the swedish and norwegian krona (exswe, exnor).[12] Note that US yield spreads are represented less than one would expect considering previous literature, with only the AAA corporate bond - federal funds spread (AAAFFM) being included in the 6 and 12-months boosting model.

2.6.2 Out-of-Sample Variable Selection

This section examines the explanatory power of variables over time throughout the hold-out sample. Figure 2.10 and 2.11 present non-zero lasso coefficients for each out-of-sample estimation for 1 and 12 months ahead forecasts. Darker values depict coefficients with higher absolute value. Following lasso coefficients over time makes

[12] It should be noted that these results do not imply causality but rather predictive power. Athey (2015), Athey and Imbens (2016) and Chernozhukov *et al.* (2015) are currently spearheading a promising research agenda aiming at combining the predictive and computational power of machine learning with causal inference that is the goal of econometric analysis.

Table 2.6: Top 10 Variables Lasso

	$h=1$		$h=3$		$h=6$		$h=12$	
1	$USTPU_1$	-0.47	$ys10_{12}$	-0.46	$ys10_1$	-0.44	$resvimf_{12}$	-0.59
2	$HOUST_1$	-0.43	$HOUST_1$	-0.36	$ys10_2$	-0.30	$ys5_0$	-0.59
3	$ys10_6$	-0.42	$ys10_4$	-0.29	$ys10_9$	-0.28	$ys5_3$	-0.27
4	$fundpur_6$	0.36	$USFIRE_1$	-0.21	$ys10_0$	-0.27	$ys10_7$	-0.22
5	$ys10_{12}$	-0.33	$ys10_5$	-0.16	$USTRADE_1$	-0.25	$NAPMSDI_9$	-0.20
6	$houst_0$	-0.28	$ys10_7$	-0.16	$USTRADE_2$	-0.23	$bondprv_5$	-0.12
7	$USFIRE_1$	-0.23	$SRVPRD_1$	-0.16	$USTRADE_3$	-0.23	$ys10_5$	-0.12
8	$fpus3_3$	0.18	$loanli_2$	0.13	$NAPMSDI_{12}$	-0.21	$bondprv_6$	-0.11
9	$bond510_0$	-0.17	BAA_1	0.10	$ys10_{11}$	-0.19	$ys5_2$	-0.11
10	$loanli_4$	0.14	$fpus3_8$	0.10	$USFIRE_1$	-0.18	$ys10_4$	-0.10

Notes: Top 10 variables with largest absolute coefficient selected by full sample lasso estimation at different forecast horizons. Upper case variable names refer to US variables, subscripts indicate lag.

Table 2.7: Top 10 Variables Boosting

	$h=1$		$h=3$		$h=6$		$h=12$	
1	$itre3_3$	34.91	$bond510_0$	51.33	$ys5_9$	35.85	$ys5_3$	42.87
2	$PAYEMS_1$	16.86	$ys5_{12}$	20.05	$iboc_2$	12.54	$exswe_{12}$	17.03
3	$fpus3_{12}$	10.78	$morres_2$	6.74	$resvimf_9$	11.05	$resvimf_{10}$	14.72
4	$PERMITS_1$	6.79	$resvimf_{12}$	5.68	$NONBOR_2$	6.70	$resvimf_3$	8.41
5	$UEMP15OV_1$	4.26	$iboc_8$	5.64	$ys10_1$	6.38	$exnor_6$	6.84
6	$houst_0$	3.17	$ys10_6$	3.84	$ys10_9$	4.14	$ys10_1$	5.78
7	$GS1_6$	2.77	$NONBOR_5$	3.31	$resvimf_{10}$	3.87	$AAAFFM_8$	4.36
8	$iboc_1$	2.71	$iboc_0$	1.76	$icbprbu_4$	3.56		
9	$ys10_6$	2.66	$USTRADE_1$	1.66	$AAAFFM_1$	2.91		
10	$NONREVSL_6$	2.29			$fpus3_{12}$	2.81		

Notes: Top 10 variables with largest relative importance selected by full sample boosting. Upper case variable names refer to US variables, subscripts indicate lag.

it possible to examine which variables were important leading indicators at specific points in time. First, note that the coefficients are not at all stable over time illustrating the structural instability of the forecasting model. The composition of the model changes specifically at the beginning and end of periods of recession providing evidence that different recessions stem from different areas of the economy. Since the end of the mid 1980's recession, 1-step ahead forecasts are dominated by US building permits and housing starts (PERMITMW, HOUST). These indicators seem to be particularly good at predicting the end of Canadian recessions. Note also that while US yield spreads were good predictors of the 1980's recession (TB6SMFFM) and the 10-year yield spread was the single best predictor of the 1990's recession, non of the yield spreads was able to predict the 2007/2008 recession.

For the 12 months ahead forecasts, the beginning of the hold-out sample estimations is characterized by spurious regression, such as the value of the Swedish Krona ($exswe_{l6}$) sharply decreasing a year before the 1976 recession. During the great moderation yield spreads become the most important predictors by far. The only other variable that is consistently featured in the model throughout the hold-out sample is Canada's reserve position with the IMF ($resvimf_{l12}$). The only notable US variable is the supplier delivery index ($NAPMSDI_9$).

The grouping of the variables also allows for a more aggregate description. Variables from the US housing and labour market complement Canadian yield spreads in short term forecasts. However, US variables are not good long term predictors of Canadian recessions that only feature financial variables and yield spreads from the Canadian economy. Contrary to popular belief, Canadian variables relating to oil and gas production are not correlated with recessions at any forecast horizon. Further note that Canadian monthly gdp does not appear as leading indicator in any of the model specifications.

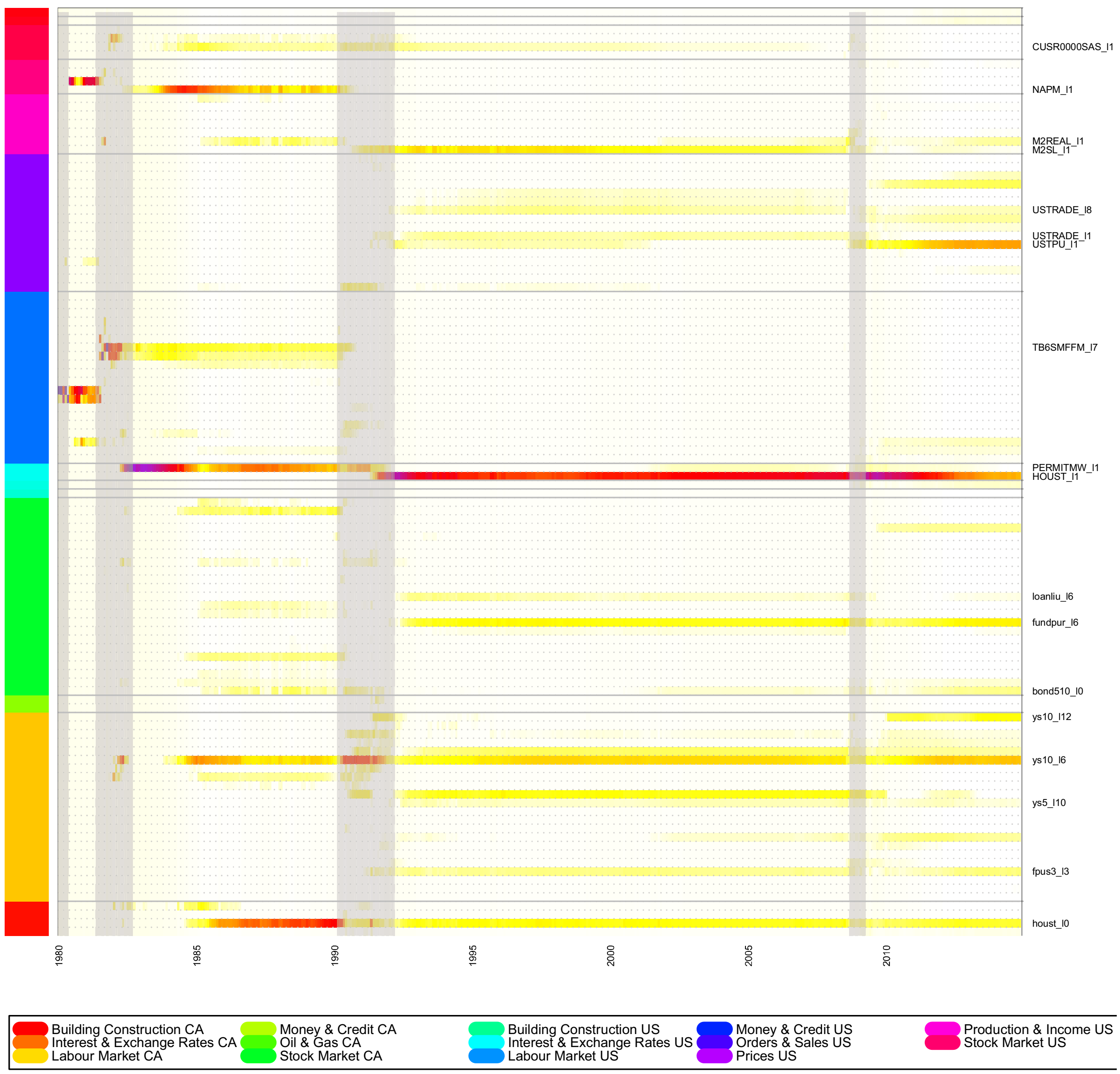

Figure 2.10: Out-of-sample variable selection using lasso at forecast horizon $h = 1$. Darker shading illustrates coefficients of higher absolute value.

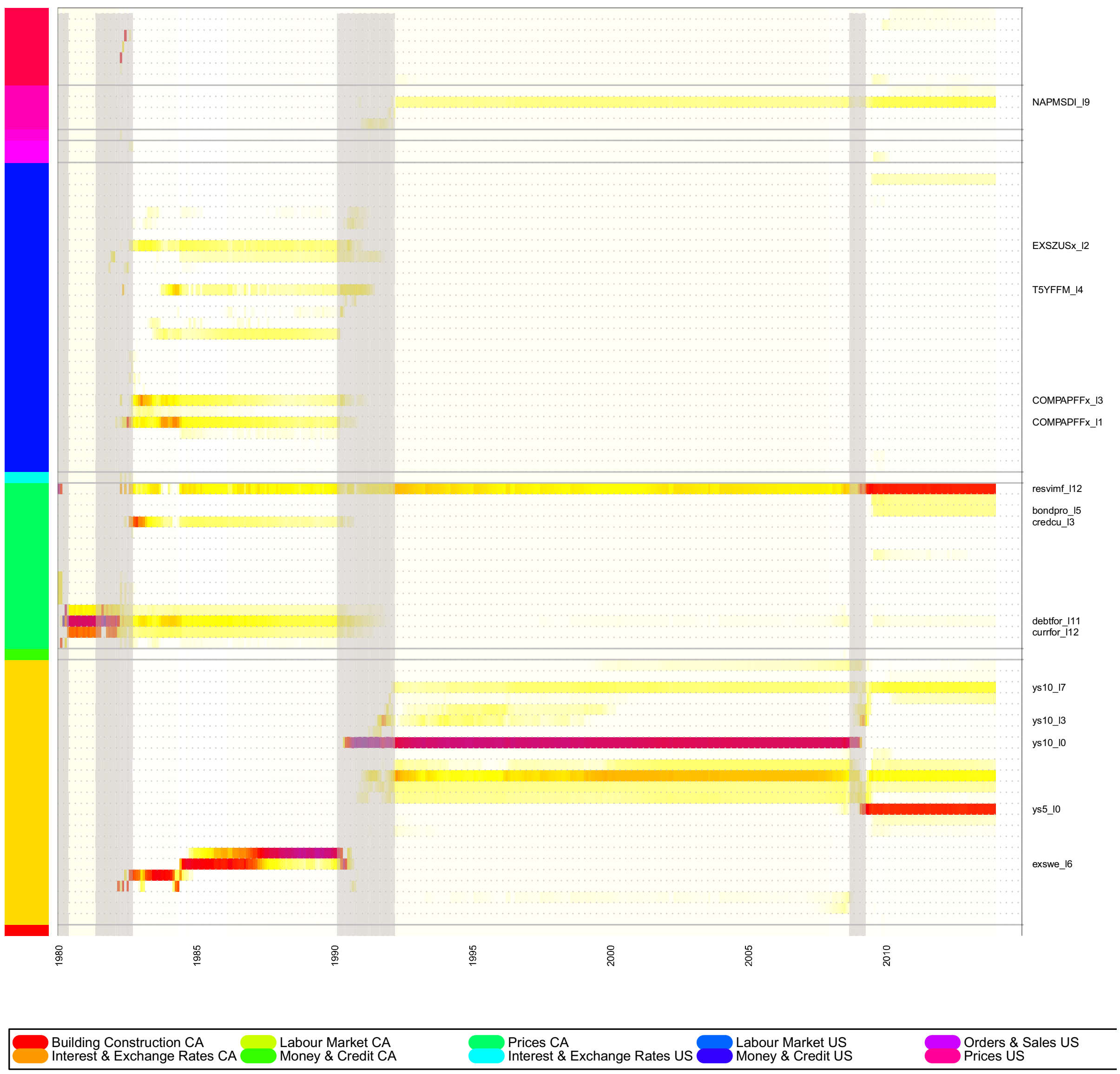

Figure 2.11: Out-of-sample variable selection using lasso at forecast horizon $h = 12$. Darker shading illustrates coefficients of higher absolute value.

2.7 Conclusion

The advent of big data opens up new opportunities for empirical research. With it come new statistical methods that are able to analyse large datasets. In this paper, I make use of two machine learning procedures that are able to select a few relevant predictors out of a dataset of over 5000 potential variables. Due to the nature of these methods, manipulating the regularization parameter can lead to a near perfect in-sample fit. Out-of-sample, the optimal value of the regularization parameter is found by cross validation to balance the bias vs. variance trade-off.

I find that lasso and boosting out-perform a naive best subset selection model. Modifications to the standard models are able to improve the out-of-sample fit slightly, but not for free. Adaptive lasso comes at the cost of a higher false positive rate, pairwise boosting comes at the cost of reduced interpretability as well as higher computational demands. Changing the validation procedure or specification, however, does not change model performance significantly.

In line with Atta-Mensah and Tkacz (1998), I find that Canadian yield spreads are important predictors at any forecast horizon. US and Canadian real activity variables add predictive power at shorter forecast horizons. In particular, while the 6 months lag of the 10-year yield spread had been the most important predictor of the 1990s recession, yield spreads lose most of their short term predictive power during the 2008/2009 recession. Instead this period is marked by a sharp decline in US housing starts. This broadens the findings of Ng and Wright (2013) by including the Canadian economy.

My results negate the finding of Bernard and Gerlach (1998), Hao and Ng (2011) and Fossati *et al.* (2017) that US yield spreads are important leading indicators of Canadian recessions. By considerably extending the number of potential predictors, the predictive power of US yield spreads appears to be picked up by financial variables from the Canadian economy. The absence of variables from the Canadian oil and gas sector is also worth discussing. One can interpret this finding as the Canadian economy being less resource dependent than previously thought. Alternatively it can be argued that recessions can coincide with periods of steep rises and steep declines in oil production or prices and therefore a quadratic transformation of these variables should be included in the estimation. I leave this exercise to future research.

Finally, while this paper uses traditional macroeconomic data, future work could easily incorporate non-traditional data such as online search results or electronic payment data. Tkacz (2013) finds some preliminary evidence in favour of the predictive power of this kind of data.

Appendix

A.1 Description of Selected Variables

Table 2.8 gives a short description of the data presented in section 2.6. This is only a small subset of the data used for analysis. The complete data set can be downloaded at http://bit.do/maxsties.

Table 2.8: Short Description of Selected Variables

Canada		United States	
ys5	5 year - 3 month yield spread	HOUST	housing starts
ys10	10 year - 3 month yield spread	PERMITS	building permits
houstsg	housing starts single units	USFIRE	employment in financial sector
exswe	exchange rate Swedish krona	USTRADE	employment in retail sector
exnor	exchange rate Norwegian krona	UEMP15OV	male unemployment rate over 15 weeks
fundpur	purchase funds	USTPU	employment in trade, transp. and util.
morres	residential mortgage rate	SRVPRD	employment in service industries
bond510	5-10 year gov. of Canada bond yield	NAPMSDI	ISM: supplier deliveries
loanli	life insurance loans	GS1	1-Year Treasury Rate
itre3	interest rate 3 month treasury bill	NONREVSL	nonrevolving Credit and busloans
fpus3	3 month US dollar forward premium	NONBOR	bank reserves
icbprbu	interest prime bus. chartered banks	AAAFFM	aaa corporate bond - fedfunds spread
bondprv	provincial bond issues	TB6SMFFM	6m treasury - fedfunds spread
resvimf	imf currency reserves	CES3000000008	avg hourly earnings : manufacturing
loanliu	life insurance loans unadjusted	REALLN	real estate loans at all com. banks
credcu	consumer credit credit unions	CLAIMSx	initial unemployment claims
currfor	foreign currency reserves	IPFUELS	industrial production: fuels
debtfor	debt held abroad	USGOOD	employment goods-prod. industries
bankacp1m	bankers' acceptance rate 1 month	BAA	moodys seasoned baa corporate bond yield
houstot	housing starts total		

A.2 Data Transformation Algorithm

1. Internal missing values are imputed linearly

2. Test for exponential growth (ols on time trend; H_0: no exponential growth)

   ⇒ if positive, apply logarithm

3. Test for heteroscedastic errors (ols of errors on moving average; H_0: no heteroscedastic)

   ⇒ if positive, apply logarithm

4. Test for seasonality (trend, cycle, seasonal decomposition; H_0: no seasonal component)

   ⇒ if positive, seasonal trend is removed

5. Test for trend stationarity (ADF; H_0: no trend stationarity)

   ⇒ if positive, linear trend is removed

6. Test for stationarity (ADF; H_0: no stationarity)

   ⇒ if negative, first difference is applied

7. Test for stationarity (again; H_0: no stationarity)

   ⇒ if negative, second difference is applied

8. All variables are standardized to have zero mean and variance of one (z-score)

A.3 Boosting Algorithm

This algorithm describes the application of non-stochastic gradient boosting using univariate regression trees employed in this paper.

Let $\hat{f}(x)$ be the function that predicts the binary reference series y.

1. **Initialization:** Set residuals equal to reference series and set boosting function equal to the constant log odds ratio

$$\hat{z}_t = y_t \text{ and } \hat{f}(x_t)^0 = \ln \frac{\sum_{t=1}^{T} y_t}{\sum_{t=1}^{T} 1 - y_t}$$

2. **Stepwise Boosting:** For $m = 1, \dots, M$, repeat

   (a) Update residuals not explained in previous step

$$\hat{z}_t^m = y_t - \hat{p}_t^{m-1}, \text{where } \hat{p_t}^{m-1} = \frac{e^{f^{m-1}(x_t)}}{1 + e^{f^{m-1}(x_t)}}$$

   (b) Find univariate regression tree that minimizes the negative binomial likelihood

$$g^m(x^j) = \arg\min \frac{-2}{T} \sum_{t=1}^{T} \left(\hat{z}_t^m g^m(x_t^j) - \ln(1 + e^{g^m(x_t^j)}) \right), j \,\epsilon\, K$$

   (c) Update $\hat{f}$ by adding a shrunken version of the new tree

$$\hat{f}^m(x) \leftarrow \hat{f}^{m-1}(x) + \mu \hat{g}^m(x)$$

3. **Average Boosted Model:**

$$\hat{F}(x) = \sum_{m=1}^{M} \mu \hat{g}^m(x) = \hat{f}^M(x)$$

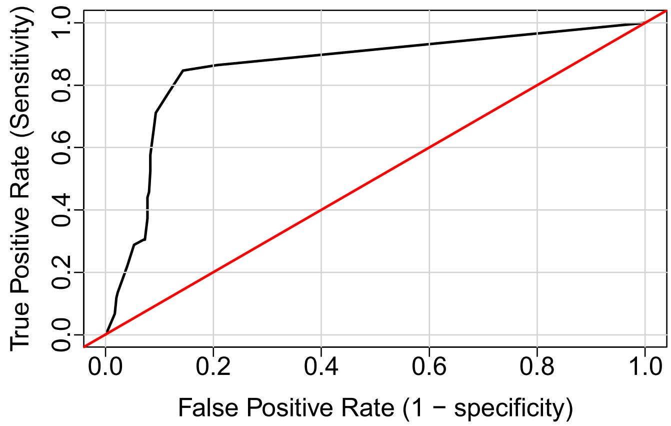

Figure 2.12: ROC curve derived from in-sample lasso estimation with $h = 0$

A.4 More on AUC and ROC

The AUC is the area under the Receiver-Operator curve that plots the true positive rate (TPR) against the false positive rate (FPR). For low values of the threshold c most estimated probabilities will be higher than the threshold even for non-recession months. The FPR will be close to one as will be the TPR. As the threshold increases, the FPR will increase faster than then TPR when estimated probabilities for non-recession months are lower than for recession months. If probabilities cannot distinguish between recession and non-recession months, FPR and TPR fall at the same rate producing a 45 line through the receiver-operator plain. The area between the actual ROC and the 45 degree line is the area under the curve (AUC), it measures how well the estimated recession probabilities are able to distinguish between recession and non-recession months. The AUC is a standard measure in the machine learning literature, I refer to Lobo *et al.* (2008) for a more detailed discussion and critique of the AUC.

Conclusion

This thesis examines the predictability of Canadian recessions using different empirical modeling techniques with an emphasis on which variables should be employed to produce recession forecasts.

Traditionally, researchers face to problem of either selecting specific variables to be included in a predictive binary regression or to aggregate the data space via factor analysis. Recently, machine learning algorithms have received increasing attention in the field of econometrics that perform variable selection and model estimation simultaneously.

The first paper in this thesis combines the two traditional approaches by comparing predictions obtained from models including only individual variables to models also allowing for factors obtained from these variables. Additionally models including only Canadian data are compared to models allowing for US data. Our findings confirm the importance of domestic yield spreads in making predictions at any forecast horizon (Atta-Mensah and Tkacz, 1998). Canadian yield spreads are best complemented with real activity indicators at short forecast horizons and with financial indicators at long forecast horizons. Including US data in the predictive regression function can improve long term forecasts marginally but is not useful at shorter forecast horizons. This result is in Bernard and Gerlach (1998) who find that US yield spreads add predictive power to Canadian recession forecasts at medium and long forecast horizons, but not at shorter ones. Additionally, we find that our estimated factor of Canadian real activity can be used to accurately predict recessions in the short term. This provides more evidence to support the finding of Castle *et al.* (2013)

who show that dynamic factors perform better than observable data at short forecast horizons. Our last result is that Bayesian model averaging assigns significant weight only to few but similar models. The best individual models receive the highest weight within the in-sample BMA forecast. Out of sample, BMA forecasts perform significantly worse than the best individual models. This provides some evidence that model averaging does not necessarily improve forecast accuracy when the reference series is binary.

The second paper increases the amount of potential data in two dimensions. First, with 134 US variables and 445 Canadian variables, a much larger set of macro series is considered; second, up to 12 lagged values of each variable are added to the dataset. The paper makes use of two machine learning procedures that are able to select a few relevant predictors out of a large dataset of over 5000 covariates. I find that lasso and boosting generally out-perform traditional methods. Modifications to the standard models are able to improve the out-of-sample fit slightly. As in the previous paper, Canadian yield spreads are important predictors at any forecast horizon. US and Canadian real activity variables add predictive power at shorter forecast horizons. In accordance with Ng and Wright (2013), who show that domestic yield spreads lost their predictive power for US recessions, I find that while the 10-year yield spread had been the most important predictor of the 1990s recession in Canada, yield spreads lose most of their short term predictive power during the 2008/2009 recession. Instead this period is marked by a sharp decline in US housing starts. By allowing for a much larger set of potential predictors, my results contradict the finding of Bernard and Gerlach (1998), Hao and Ng (2011), that US yield spreads are important leading indicators of Canadian recessions. The predictive power of US yield spreads found in previous papers appears to be picked up by financial variables from the Canadian economy.

Another interesting distinction between the two paper in this thesis is that the variables picked at different points in time in the first paper were generally very similar for a specific forecast horizon. However, once a larger amount of predictors is available, the selection of variables over time becomes much more inconsistent. This points to the biggest limitation in any recession forecasting exercise: Any forecast is always backwards looking. Since only previous information can be utilized to produce forecasts, recessions that originate from a sector of the economy that previously had not been associated with periods of recession, like the US housing market during 2007/2008, are extremely hard if not impossible to predict.

Bibliography

Arlot, S., Celisse, A. *et al.* (2010) A survey of cross-validation procedures for model selection, *Statistics surveys*, **4**, 40–79.

Athey, S. (2015) Machine learning and causal inference for policy evaluation, in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 5–6.

Athey, S. and Imbens, G. (2016) Recursive partitioning for heterogeneous causal effects, *Proceedings of the National Academy of Sciences*, **113**, 7353–7360.

Athey, S. and Imbens, G. (2017) The state of applied econometrics - causality and policy evaluation, *Journal of Economic Perspective*, **31**, 332.

Atta-Mensah, J. and Tkacz, G. (1998) Predicting Canadian recessions using financial variables: A probit approach, Working Paper 98-5, Bank of Canada.

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

Beaton, K., Lalonde, R. and Snudden, S. (2014) The propagation of US shocks to Canada: Understanding the role of real financial linkages, *Canadian Journal of Economics/Revue canadienne d'économique*, **47**, 466–493.

Berge, T. J. (2015) Predicting recessions with leading indicators: Model averaging and selection over the business cycle, *Journal of Forecasting*, **34**, 455–471.

Berge, T. J. and Jorda, O. (2011) Evaluating the classification of economic activity in recessions and expansions, *American Economic Journal: Macroeconomics*, **3**, 246–277.

Bernard, H. and Gerlach, S. (1998) Does the term structure predict recessions? the international evidence, *International Journal of Finance & Economics*, **3**, 195–215.

Bragoli, D. and Modugno, M. (2016) A nowcasting model for Canada: Do US variables matter?, FEDS Working Paper 2016-036, Board of Governors of the Federal Reserve System.

Bühlmann, P. and Yu, B. (2003) Boosting with the L2 loss: regression and classification, *Journal of the American Statistical Association*, **98**, 324–339.

Burns, A. F. and Mitchell, W. C. (1938) Statistical indicators of cyclical revivals, in *Business Cycle Indicators* (Ed.) G. H. Moore, Princeton University Press, vol. 1, pp. 184–260, reprinted 1961 edn.

Business Cycle Council (2012) C.D. Howe institute business cycle council issues authoritative dates for the 2008/2009 recession.

Camacho, M., Perez-Quiros, G. and Poncela, P. (2015) Extracting nonlinear signals from several economic indicators, *Journal of Applied Econometrics*, **30**, 1073–1089.

Castle, J. L., Clements, M. P. and Hendry, D. F. (2013) Forecasting by factors, by variables, by both or neither?, *Journal of Econometrics*, **177**, 305–319.

Chauvet, M. (1998) An econometric characterization of business cycle dynamics with factor structure and regime switches, *International Economic Review*, **39**, 969–996.

Chauvet, M. and Piger, J. (2008) A comparison of the real-time performance of business cycle dating methods, *Journal of Business & Economic Statistics*, **26**, 42–49.

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

Chauvet, M. and Potter, S. (2010) Business cycle monitoring with structural changes, *International Journal of Forecasting*, **26**, 777–793.

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

Chernozhukov, V., Hansen, C. and Spindler, M. (2015) Valid post-selection and post-regularization inference: An elementary, general approach, *Annual Review of Economics*, **7**, 649–688.

Cramer, J. S. (1999) Predictive performance of the binary logit model in unbalanced samples, *Journal of the Royal Statistical Society: Series D (The Statistician)*, **48**, 85–94.

Cross, P. and Bergevin, P. (2012) Turning points: Business cycles in Canada since 1926, Commentary 366, C.D. Howe Institute.

De Nicolò, G. and Lucchetta, M. (2016) Forecasting tail risks, *Journal of Applied Econometrics*.

Diebold, F. X. and Rudebusch, G. D. (1996) Measuring business cycles: A modern perspective, *Review of Economics and Statistics*, **78**, 66–77.

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

Estrella, A. and Hardouvelis, G. A. (1991) The term structure as a predictor of real economic activity, *The journal of Finance*, **46**, 555–576.

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

Faust, J., Gilchrist, S., Wright, J. H. and Zakrajšsek, E. (1996) Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach, *Review of Economics and Statistics*, **95**, 1501–1519.

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

Fossati, S. (2015) Forecasting US recessions with macro factors, *Applied Economics*, **47**, 5726–5738.

Fossati, S. (2016) Dating US business cycles with macro factors, *Studies in Nonlinear Dynamics & Econometrics*, **20**, 529–547.

Fossati, S., Sekkel, R. and Sties, M. (2017) Forecasting recessions in Canada, Working paper.

Freund, Y., Schapire, R. E. *et al.* (1996) Experiments with a new boosting algorithm, in *ICML*, vol. 96, pp. 148–156.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software*, **33**, 1.

Friedman, J., Hastie, T., Tibshirani, R. *et al.* (2000) Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors), *The annals of statistics*, **28**, 337–407.

Friedman, J. H. (2001) Greedy function approximation: A gradient boosting machine, *Annals of statistics*, pp. 1189–1232.

Friedman, J. H. (2002) Stochastic gradient boosting, *Computational Statistics & Data Analysis*, **38**, 367–378.

Gaudreault, C., Lamy, R. and Liu, Y. (2003) New coincident, leading and recession indexes for the Canadian economy: An application of the Stock and Watson methodology, Working Paper 2003-12, Department of Finance Canada, Economic and Fiscal Policy Branch.

Gosselin, M.-A. and Tkacz, G. (2010) Using dynamic factor models to forecast Canadian inflation: the role of US variables, *Applied Economics Letters*, **17**, 15–18.

Groen, J. J., Paap, R. and Ravazzolo, F. (2013) Real-time inflation forecasting in a changing world, *Journal of Business & Economic Statistics*, **31**, 29–44.

Guérin, P. and Leiva-Leon, D. (2014) Model averaging in markov-switching models: Predicting national recessions with regional data, Working Paper 60250, MPRA.

Hao, L. and Ng, E. C. (2011) Predicting Canadian recessions using dynamic probit modelling approaches, *Canadian Journal of Economics/Revue canadienne d'économique*, **44**, 1297–1330.

Hoogerheide, L., Kleijn, R., Ravazzolo, F., Van Dijk, H. K. and Verbeek, M. (2010) Forecast accuracy and economic gains from bayesian model averaging using time-varying weights, *Journal of Forecasting*, **29**, 251–269.

Horowitz, J. L. (2015) Variable selection and estimation in high-dimensional models, *Canadian Journal of Economics/Revue canadienne d'économique*, **48**, 389–407.

Katayama, M. (2010) Improving recession probability forecasts in the US economy.

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

Kim, C.-J. and Nelson, C. R. (1998) Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching, *Review of Economics and Statistics*, **80**, 188–201.

Kim, C.-J. and Nelson, C. R. (1999) *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, MIT Press.

Li, J. and Chen, W. (2014) Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models, *International Journal of Forecasting*, **30**, 996–1015.

Lobo, J. M., Jiménez-Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models, **17**, 145–151.

Lokhorst, J. (1999) The lasso and generalised linear models, *Honors Project, The University of Adelaide, Australia*.

Ludvigson, S. C. and Ng, S. (2010) A factor analysis of bond risk premia, in *Handbook of Empirical Economics and Finance* (Eds.) A. Uhla and D. E. A. Giles, Chapman and Hall, Boca Raton, chap. 12, pp. 313–372.

Mayr, A., Binder, H., Gefeller, O., Schmid, M. *et al.* (2014) The evolution of boosting algorithms, *Methods of Information in Medicine*, **53**, 419–427.

McCracken, M. and Ng, S. (2014) FRED-MD: A monthly database for macroeconomic research, Working Paper 2015-012A, Federal Reserve Bank of St. Louis, http://www.columbia.edu/sn2294/papers/freddata.pdf.

Ng, E. C. (2012) Forecasting US recessions with various risk factors and dynamic probit models, *Journal of Macroeconomics*, **34**, 112–125.

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

Ng, S. and Wright, J. H. (2013) Facts and challenges from the recession for forecasting and macroeconomic modeling, *Journal of Economic Literature*, **51**, 1120–1154.

Nyberg, H. (2010) Dynamic probit models and financial variables in recession forecasting, *Journal of Forecasting*, **29**, 215–230.

Raftery, A. E. (1995) Bayesian model selection in social research, *Sociological Methodology*, **25**, 111–164.

Sampson, J. N., Chatterjee, N., Carroll, R. J. and Müller, S. (2013) Controlling the local false discovery rate in the adaptive lasso, *Biostatistics*, **14**, 653–666.

Schapire, R. E. (1990) The strength of weak learnability, *Machine learning*, **5**, 197–227.

Shevade, S. K. and Keerthi, S. S. (2003) A simple and efficient algorithm for gene selection using sparse logistic regression, *Bioinformatics*, **19**, 2246–2253.

Sties, M. (2017) Forecasting Canadian recessions: Making use of supervised machine learning.

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

Stock, J. H. and Watson, M. W. (1992) New indexes of coincident and leading economic indicators, in *Leading economic indicators: New approaches and forecasting records* (Eds.) K. Lahiri and G. H. Moore, Cambridge University Press, chap. 4, pp. 63 – 90.

Stock, J. H. and Watson, M. W. (1993) A procedure for predicting recessions with leading indicators: econometric issues and recent experience, in *Business cycles, indicators and forecasting* (Eds.) J. H. Stock and M. W. Watson, University of Chicago Press, pp. 95–156.

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

Stock, J. H. and Watson, M. W. (2006) Forecasting with many predictors, *Handbook of economic forecasting*, **1**, 515–554.

Stock, J. H. and Watson, M. W. (2016) Factor models and structural vector autoregressions in macroeconomics, in *Handbook of Macroeconomics* (Eds.) J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 8, pp. 415–525.

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

Tkacz, G. (2013) Predicting recessions in real-time: Mining Google trends and electronic payments data for clues, Commentary 387, C.D. Howe Institute.

Varian, H. R. (2014) Big data: New tricks for econometrics, *The Journal of Economic Perspectives*, **28**, 3–27.

Wright, J. H. (2008) Bayesian model averaging and exchange rate forecasts, *Journal of Econometrics*, **146**, 329–341.

Zou, H. (2006) The adaptive lasso and its oracle properties, *Journal of the American statistical association*, **101**, 1418–1429.

Zou, H. and Hastie, T. (2005) Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 301–320.