Journal of Forecasting, *J. Forecast.* **34**, 455–471 (2015)
Published online 13 May 2015 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2345

Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle

TRAVIS J. BERGE*

Board of Governors of the Federal Reserve System, Washington, DC, USA

ABSTRACT

Four methods of model selection—equally weighted forecasts, Bayesian model-averaged forecasts, and two models produced by the machine-learning algorithm boosting—are applied to the problem of predicting business cycle turning points with a set of common macroeconomic variables. The methods address a fundamental problem faced by forecasters: the most useful model is simple but makes use of all relevant indicators. The results indicate that successful models of recession condition on different economic indicators at different forecast horizons. Predictors that describe real economic activity provide the clearest signal of recession at very short horizons. In contrast, signals from housing and financial markets produce the best forecasts at longer forecast horizons. A real-time forecast experiment explores the predictability of the 2001 and 2007 recessions. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS business cycle turning points; variable selection; boosting; Bayesian model averaging; probabilistic forecasts

INTRODUCTION

A common view of the behavior of modern economies is that economic output oscillates around a trend rate of growth, alternately experiencing phases of expansion and recession. Economic activity grows during expansions, increasing standards of living. However, these periods of expansion are followed by sudden and rapid declines in activity, observed across a large number of sectors in the economy. Since the reduction in activity is broad based, these recessions suggest that households face a higher unemployment rate and stagnant wage growth. A growing literature documents other pernicious effects of recession, ranging from long-lasting negative earnings effects to negative consequences for individual health and educational outcomes.[1] For businesses, recession decreases profitable economic opportunities and investment, and may result in the reduction of payroll employment.

Given these observations, the enthusiasm with which households, businesses and policymakers attempt to infer the current and future states of the economy comes as no surprise. Classifying economic variables into variables that lead, are coincident to and lag economic downturns is a long-lived tradition in economic research, going back to at least Burns and Mitchell (1946). The financial press and economic practitioners carefully parse a wide range of economic and financial indicators. While many of these indicators contain information useful for the identification of the current or future states of the economy, many others likely do not.

In the United States, the business cycle dating committee of the NBER produces the gold-standard recession chronology. The committee does not rely on a particular model to define recession. Instead, a range of economic indicators are used to determine peaks and troughs of economic activity, including measures of production, income, employment and sales. The weights given to each indicator are explicitly ambiguous and may change over time.[2] That there is no mechanical model to be applied complicates any forecasting exercise and places model selection at the heart of the problem. Further, Ng and Wright (2013) argue that the behavior of many economic indicators across the business cycle has changed. Similarly, Chauvet and Potter (2002, 2005) find evidence of breaks in a common predictor of recession: the slope of the yield curve. Recognizing that GDP is an incomplete indicator of aggregate economic activity, there is a related research agenda focused on producing a statistic that summarizes the cyclical position of the economy, as in Stock and Watson (1999)—implemented as the CFNAI at the Federal Reserve Bank of Chicago—Aruoba *et al.* (2009) and others. Hamilton (2011) argues that while factor models are accurate descriptions of the state of the economy in-sample, their usefulness as forecasting tools is limited if they incorporate misleading indicators, or if the behavior of the incorporated indicators has changed over time.

* Correspondence to: Travis J. Berge, Board of Governors of the Federal Reserve System, 20th St and C St NW, Washington, DC 20431, USA. E-mail: travis.j.berge@frb.gov

[1] See, among others, Ruhm (1991); Oreopoulos *et al.* (2006); Sullivan and von Wachter (2009); Stevens and Schaller (2011). While Ruhm (2000) has documented the pro-cyclicality of the average mortality rate, Stevens *et al.* (2011) argue that cyclical fluctuations in mortality are driven largely by the mortality rate of the elderly, which itself depends on health care quality, and is sensitive to the state of the labor market.

[2] See www.nber.org/cycles/sept2010.html.

This paper evaluates the predictive content of a number of commonly followed macroeconomic variables, focusing on the evaluation of probabilistic forecasts of recession. Equally weighted forecasts and Bayesian model-averaged forecasts are produced and compared to predictions from model selection algorithm boosting. Boosting endogenously selects covariates to be included in a forecast model. While boosting originated in the machine-learning literature focusing primarily on problems of classification, it is increasingly applied to empirical problems in other disciplines, including economics.[3] The method can be specified non-parametrically—though the approach taken here is akin to a logistic regression—is highly efficient, and resistant to overfitting.

The results indicate that many indicators considered do indeed contain information that can be exploited to produce worthwhile forecasts of future states of the economy. Real economic variables, especially indicators of the labor market, most accurately describe the current state of the economy. The results also point to a strong relationship between the bond market and real economic activity, albeit one that occurs with a lag. The yield curve is shown yet again to be a predictor of future turning points, with the caveat that its predictive power is limited to forecasts made at particular horizons, and that it did not provide very strong signals of the 2001 or 2007 recessions. Other variables contain information useful for forecasting into the medium term, such as corporate bond spreads and indicators of the housing market. Of the methods considered, BMA and the boosting algorithm each produce useful probabilistic forecasts of recession. In contrast, unweighted model averages do not forecast well, because they cannot discriminate disparate signals of recession. Finally, the model selection schemes find considerable time variation in the predictors most relevant to the forecasting problem.

EMPIRICAL SETUP

Let y_t denote the state of the business cycle as determined by the NBER, where $y_t = 1$ denotes that month t is an NBER-defined recession and $y_t = 0$ is an expansion. The logistic model assumes that the log-odds ratio is a function of a vector of covariates, x_t, so that

$$\log \frac{P[y_t = 1|x_{t-h-1}]}{P[y_t = 0|x_{t-h-1}]} \equiv y_t = f(x_{t-h-1}, \Theta) \tag{1}$$

where x is a $(K + 1) \times 1$ vector of K observables plus a constant, Θ denotes model parameters and the subscript h denotes the forecast horizon.[4] Under the assumption of linearity, i.e. $f(x_{t-h-1}, \Theta) = x'_{t-h-1}\beta$, the likelihood of the logistic is standard and parameter estimates can be obtained via maximum likelihood, for example.

A large literature has used models similar to equation (1) to relate the state of the business cycle to forward-looking economic indicators, primarily the term structure of interest rates. Estrella and Mishkin (1998), Wright (2006) and Rudebusch and Williams (2009), among others, focus on simple limited dependent models that condition on the slope of the yield curve. Each argues that simple specifications predict turning points quite well. Dueker (2005), Chauvet and Potter (2005) and Kauppi and Saikkonen (2008) argue that dynamic specifications improve on forecast ability, and one shortcoming of the methods used here is that they are static. However, as Marcellino *et al.* (2006) point out, whether iterated forecasts from a dynamic model dominate direct forecasts is an empirical matter; iterated forecasts are more efficient only if the model is correctly specified.

While the yield curve is the best-known leading indicator of economic downturns, there are many reasons why conditioning on additional economic indicators is likely to improve forecast ability. First, the yield curve does not forecast well at all forecast horizons. The yield curve may also be an unstable predictor of recessions as risk and term premia are time varying, and may behave differently in a low-interest rate environment. To the extent that the yield curve is a summary statistic of market expectations for the future path of short-term interest rates, changes to the monetary reaction function may muddle simple limited dependent models using the yield curve. Finally, since shocks to the real economy alter the probability of future recession, inclusion of other variables that directly measure the real economy likely improve forecast ability.

One natural method for combining information from a wide range of sources is to use a factor model, as in Stock and Watson (1989, 1999). A number of papers identify business cycle turning points with factor models. Chauvet (1998), Chauvet and Hamilton (2006), Chauvet and Piger (2008) and Chauvet and Senyuz (2012) relate the state of the economy to dynamic factors extracted from macroeconomic or financial indicators. Chen *et al.* (2011) and Fossati (2011) use factor models within a probit setup to forecast recessions. However, there is no clear reason to believe that the unobserved factor that best captures the cross-sectional variation in a panel of data (for example) will forecast future states of the economy (Bai and Ng, 2009). Indeed, the handful of papers focused on forecasting turning points with factor models typically have limited success as the forecast horizon grows.

[3] See, for example, Bai and Ng (2009); Khandani *et al.* (2010); Berge (2014); Ng (2014).

[4] The subscript $t - h - 1$ indicates that many of the indicators included in the model search have a publication lag of 2–3 weeks. In the forecast experiment, the forecaster waits until each indicator is available, the third week of the month, then produces forecasts. A 'nowcast', where $h = 0$, is thus a forecast of the state of the economy the previous month.

Finally, a handful of papers have applied methods similar to those used here to identify or forecast recessions. Owyang *et al.* (2013) use a Bayesian model-averaging methodology to show that state-level employment data identify turning points in the national economy. Giusto and Piger (2013) use machine-learning algorithms to evaluate how quickly one can call a recession in real time. Their focus is minimizing the amount of time between the realization of a business cycle turning point and its identification. Independent from this paper, Ng (2014) focuses on forecasting recessions with a boosting algorithm, albeit one that is differently specified from the one applied here. She finds the forecast ability of the algorithm is mixed and argues that time variation in the best forecasting model complicates the problem. This paper is complementary to Ng; the two papers use slightly different algorithms and predictors but reach broadly similar conclusions. The current paper takes the additional step of comparing the boosting algorithm to other common methods of model selection.

Averaging model forecasts

Model averaging has long been recognized as a useful method of combining information from a given set of models. Previous applications have shown that model averaging tends to improve forecast accuracy, either because the combination incorporates information from overlapping information sets (Bates and Granger, 1969), or because the combination alleviates model misspecification (Hendry and Clements, 2004; Stock and Watson, 2004; Timmermann, 2006). Each of a set of M models is estimated to produce a forecast of y_t, resulting in $\{\hat{y}_{1t}, \hat{y}_{2t}, \ldots, \hat{y}_{Mt}\}$. The model-averaging problem is to find weights w to produce a single forecast $\hat{y}_t^C = C(\hat{y}_{1t}, \ldots, \hat{y}_{Mt}, w_1, \ldots, w_M)$. In principle, because forecasts impact the actions of economic agents, the weights could reflect the utility of decision makers (Elliott and Lieli, 2013). In the current application, the trade-off between true and false positives or true and false negatives for recession forecasts is unclear. Thus, instead of focusing on a loss function, two empirical weighting schemes are utilized. The first weights different forecasting models equally. In the second scheme, model fit is used to produce weights for the combined recession forecast.

The equally weighted forecast assumes that there are K covariates. Each model is a univariate forecasting model, so that the number of models is equal to K. Let $\hat{y}_{it}$ denote the forecast for date t of model i. The equally weighted forecast is

$$\hat{y}_t^{\mathrm{EW}} = \frac{1}{K}\sum_{i=1}^{K} \hat{y}_{it} \tag{2}$$

The second weighting scheme, Bayesian model averaging (BMA), allows a richer set of models to be considered. BMA allows each possible combination of covariates, a total of $M = 2^K$ models. Let each model M_i be parametrized by Θ_i. The Bayesian model average forecast is the probability-weighted sum of the model-specific forecasts:

$$\hat{y}_t^{\mathrm{BMA}} = \sum_{i=1}^{M} \hat{y}_{it}\mathrm{Pr}(M_i|D_{t-h-1}) \tag{3}$$

where $\hat{y}_{it}$ is the forecast from model i and $Pr(M_i|D_{t-h-1})$ denotes the posterior probability of model i conditional on the data available at the time the forecast is made. The posterior probability of model i is proportional to that model's likelihood multiplied by its prior. Let $P(M_i)$ denote the prior that model i is true. The researcher observes the data D then computes the posterior probability for each model i:

$$\mathrm{Pr}(M_i|D) = \frac{\mathrm{Pr}(D|M_i)\mathrm{Pr}(M_i)}{\sum_{j=1}^{M} P(D|M_j)P(M_j)} \tag{4}$$

where $\mathrm{Pr}(D|M_i) = \int \mathrm{Pr}(D|\theta_i, M_i)\mathrm{Pr}(\theta_i|M_i)\partial\theta_i$. Directly implementing equation (4) requires the calculation of a marginal likelihood, which is demanding because the average is taken over the prior distribution for all model parameters. However, the Bayesian information criterion (BIC) is a consistent estimate of the marginal likelihood of a model (Raftery, 1995). This estimate of the marginal likelihood is commonly used in applied work, and is advantageous since it requires only a maximum likelihood estimate (see, for example, Sala-i-Martin *et al.*, 2004; Brock *et al.*, 2007; Morley and Piger, 2012). Each model is assumed equally likely a priori, so that model posterior probabilities are calculated as model fit relative to the fit of all models:

$$\mathrm{Pr}(M_i|D) = \frac{\exp(\widehat{\mathrm{BIC}}_i)}{\sum_{i=1}^{M}\exp(\widehat{\mathrm{BIC}}_i)}$$

Model selection via the boosting algorithm

Boosting estimates the function $F : \mathbb{R}^K \to \mathbb{R}$ that minimizes the expected loss $\mathcal{L}(y, F)$ of some loss function:

$$\hat{F}(x) \equiv \underset{F(x)}{\arg\min}\, E\left[\mathcal{L}(y, F(x))\right] \tag{5}$$

The setup encompasses many different regression-like problems. Specifying $\mathcal{L}(y, F(x))$ as squared-error loss results in a problem analogous to least-squares regression, but loss functions specific to limited dependent variables, count data or duration data are also easily implemented. In the implementation below, the loss function is specified to be one-half times the negative of the Bernoulli likelihood function. With this loss function the algorithm is referred to as *logitboost*. Recall the parametric logistic model in equation (1), which specified the log-odds ratio as a linear function of a set of covariates. Boosting can be thought of as a non-parametric extension of that model, and computationally the application below is a stage-wise additive logistic model.

The algorithm solves for $F(x)$ in a step-wise manner. $F(x)$—known as the *strong learner*—is a linear combination of simple models, known as *weak learners*:

$$F_I(x) = \sum_{i=1}^{I} \rho_i f_i(x, \theta_i) \tag{6}$$

where i indicates iterations of the algorithm, ρ_i is a weight for the weak learner selected at each iteration and θ_i is a parameter vector.

The algorithm that minimizes the empirical counterpart to equation (5) is due to Friedman (2001), implemented using the R package from Buhlmann and Hothorn (2007) and summarized as follows. After initializing, step 3 of the algorithm fits each weak learner to the negative gradient of the loss function given the current estimate of the strong learner. Step 4 searches across the weak learners to choose the one that most quickly descends the function space. Step 5 iterates on 2–4 until a terminal iteration I.

Functional gradient descent:

1. *Initialize the model.* Choose a functional form for each weak learner, $f^{(k)}, k = 1, \ldots, K$. Typically, each covariate k receives its own functional form, which need not be identical across variables.

   Let i denote iterations of the algorithm, and set $i = 0$. Initialize the strong learner F_0. It is common to set F_0 equal to the constant c that minimizes the empirical loss.
2. *Increase i by 1.*
3. *Projection.* Compute the negative gradient of the loss function evaluated at the current estimate of F, $\hat{F}_{i-1}$. This produces

$$\mathbf{u_i} \equiv \{u_{i,t}\}_{t=1,\ldots,T} = -\frac{\partial \mathcal{L}(y_t, F)}{\partial F}|_{F=\hat{F}_{i-1}(x_t)}, t = 1, \ldots, T$$

   Fit each weak learner to the current negative gradient vector $\mathbf{u_i}$.
4. *Update F_i.* Let $\hat{f}_i^{(\kappa)}$ denote the weak learner with the smallest residual sum of squares among the K weak learners. Update the estimate of F by adding the weak learner κ to the estimate of F_{i-1}:

$$\hat{F}_i = \hat{F}_{i-1} + \rho \hat{f}_i^{\kappa}$$

   It is common to simply use a constant but small shrinkage factor, ρ, although an additional minimization problem can be solved to find the best step size.
5. *Iterate.* Iterate on steps 2–4 until $i = I$.

Each weak learner must be specified *ex ante*. Ng (2014) uses the above setup to model the log-odds ratio as a non-parametric function of predictors, wherein each covariate is modeled as a decision tree with only one node (a 'stump'). The initial approach taken here is analogous to a logistic model in that the log-odds is assumed to be linear in each covariate. In the algorithm, at each iteration, the covariate with the best linear fit at that iteration is added to the model. After iterating, the final model becomes

$$\begin{aligned}
\hat{F}_I(x) &= \sum_{i=1}^{I} \rho_i \hat{f}_i(x) \\
\hat{f}_i(x) &= f^{\kappa}(x) \\
\kappa &= \underset{k}{\arg\min} \sum_{t=1}^{T} \left(u_t - \hat{f}^k(x_t)\right)^2
\end{aligned} \tag{7}$$

Table I. Variables included in forecasting models

Variable	Definition	Transformation
Interest rates and interest rate spreads		
Level of yield curve	Average of 3-month, 2- and 10-year yields	—
Slope of yield curve	10-year less 3-month yield	—
Curvature of yield curve	2 × 2-year minus sum of 3-month and 10-year yield	—
TED spread	3-month ED less 3-month Treasury yield	—
BAA corporate spread	BAA less 10-year Treasury yield	—
AAA corporate spread	AAA less 10-year Treasury yield	—
Other financial variables		
Change in stock index	S&P 500 Index	3-month log difference
Money growth	M2	3-month log difference
Real money growth	M2 deflated by CPI	3-month log difference
US dollar	Trade-weighted dollar	3-month log difference
VIX	VIX from CBOE and extended following Bloom	—
Macroeconomic indicators		
Output	Industrial production (s.a.)	3-month log difference
Income	Real personal income (s.a.)	3-month log difference
Housing permits	—	3-month log difference
Total employment	Payroll employment	3-month log difference
Initial claims	4-week moving average (s.a.)	3-month log difference
Weekly hours, manufacturing	—	3-month log difference
Civilian unemployment rate	—	3-month change
Purchasing managers index	—	3-month log difference

The parameters $\{\rho_i\}_{i=1}^{I}$ and I jointly determine the size of the model and model fit. The step-size ρ_i is fixed to equal 0.1. The number of iterations, I, minimizes the Schwarz information criterion; i.e. $I \equiv \arg\min_i \mathrm{BIC}(i)$.[5] Since the function is estimated in a stage-wise manner, some covariates will be selected many times through the course of the algorithm, whereas others may never enter the final model.

In a standard logistic model, the log-odds ratio is assumed to depend linearly on the covariates. However, this functional form is used only because it is convenient, and there are good reasons to believe that a nonlinear specification is appropriate for the current application. It is straightforward to introduce nonlinearity by altering the specification of the weak learners $f(x)$. The nonlinear specification in the application below assumes the smoothing splines of Eilers and Marx (1996). Each covariate k minimizes the penalized sum of squared error:

$$\mathrm{PSSE}(f^k, \lambda) = \sum_{t=1}^{T} \left[u_t - f^k(x_t)\right]^2 + \lambda \int [f^{k''}(z)]^2 \mathrm{d}z \tag{8}$$

where the smoothing parameter λ determines the magnitude of the penalty for functions with a large second derivative. Buhlmann and Yu (2003) recommend setting $\lambda = 4$, which is the value used here. As with the linear case, at each iteration the covariate that best minimizes the empirical loss is included in the forecast model, and I is selected to minimize the BIC of the final model.

DATA AND EVALUATION

The analysis includes the commonly followed financial and macroeconomic indicators listed in Tables I and AI. These variables describe many different sectors of the economy, with enough indicators to make the selection problem meaningful. Several variables describe the real economy. Industrial production serves as a proxy for output. Housing permits proxy for the housing market. Labor market indicators include the 4-week moving average of initial claims for unemployment insurance, a measure of hours worked and the 3-month change in the unemployment rate. The purchasing managers index, a commonly followed leading indicator, is also included in the model. A variety of financial indicators are also included. The level, slope and curvature of the yield curve are constructed using monthly averages of the daily yields of zero-coupon 3-month, 2-year and 10-year yields compiled by Gurkaynak *et al.* (2007). The TED spread and two corporate bond spreads measure credit risk. Money growth rates, a

[5] Specifically, $\mathrm{BIC}(i) = -2\sum_{t=1}^{T} [y_t \log(\hat{p}_{it}) + (1-y_t) log(1-\hat{p}_{it})] + \log(T) \times \mathrm{df}_i$, where $\hat{p}_{it}$ denotes the probability of recession estimated for period t by the model produced at iteration i, T is the number of observations and df_i is the degrees of freedom of the model at iteration i, measured by the trace of the hat-matrix of the model at that iteration (Buhlmann and Hothorn, 2007).

Table II. In-sample statistics for univariate forecasts for each indicator. Sample size 1972:3–2013:12. The first row presents the pseudo-R^2, the second row the quadratic probability score, the third row the accuracy, and the final row gives the AUC. Bold entries indicate best-performing value for that forecast horizon

		$h=0$	$h=6$	$h=12$	$h=18$	$h=24$		$h=0$	$h=6$	$h=12$	$h=18$	$h=24$
Level	Psd.R^2	0.04	0.06	0.05	0.04	0.01	Dollar	0.01	0.01	0.00	0.00	0.00
	QPS	0.13	0.12	0.12	0.12	0.11		0.13	0.13	0.13	0.12	0.11
	ACC	0.84	0.83	0.84	**0.85**	**0.86**		0.84	0.83	0.84	**0.85**	**0.86**
	AUC	0.62	0.66	0.66	0.65	0.60		0.57	0.55	0.54	0.55	0.53
Slope	Psd.R^2	0.03	**0.23**	**0.27**	0.20	0.11	IP	0.32	0.04	0.00	0.00	0.00
	QPS	0.13	**0.10**	**0.09**	0.10	0.10		0.08	0.13	0.13	0.12	0.11
	ACC	0.84	**0.85**	**0.85**	**0.85**	**0.86**		0.89	0.82	0.84	**0.85**	**0.86**
	AUC	0.63	**0.83**	**0.89**	0.86	0.78		0.88	0.71	0.59	0.48	0.53
Curve	Psd.R^2	0.01	0.00	0.01	0.02	0.02	Real p.i.	0.14	0.03	0.00	0.00	0.00
	QPS	0.13	0.14	0.13	0.12	0.11		0.11	0.13	0.13	0.12	0.11
	ACC	0.84	0.83	0.84	**0.85**	**0.86**		0.85	0.82	0.84	**0.85**	**0.86**
	AUC	0.53	0.50	0.54	0.61	0.60		0.78	0.65	0.50	0.52	0.55
AAA-10Y	Psd.R^2	0.02	0.14	0.21	**0.23**	**0.13**	New permits	0.15	0.13	0.02	0.02	0.00
	QPS	0.13	0.11	0.10	**0.09**	**0.10**		0.1	0.11	0.12	0.12	0.11
	ACC	0.84	0.84	**0.85**	0.84	**0.86**		0.87	**0.85**	0.84	**0.85**	**0.86**
	AUC	0.60	0.78	0.85	**0.88**	**0.80**		0.76	0.76	0.68	0.67	0.59
BAA-10Y	Psd.R^2	0.00	0.10	0.19	0.22	**0.13**	ISM PMI	0.10	0.04	0.00	0.00	0.00
	QPS	0.13	0.12	0.10	0.10	**0.10**		0.11	0.13	0.13	0.12	0.11
	ACC	0.84	0.82	0.84	0.83	0.85		0.86	0.82	0.84	**0.85**	**0.86**
	AUC	0.52	0.75	0.83	0.86	0.79		0.69	0.68	0.62	0.58	0.50
Ted spread	Psd.R^2	0.27	0.19	0.07	0.04	0.01	Weekly hrs	0.11	0.02	0.01	0.00	0.00
	QPS	0.09	0.11	0.12	0.12	0.11		0.12	0.13	0.13	0.12	0.11
	ACC	0.88	0.84	0.83	**0.85**	**0.86**		0.84	0.83	0.84	**0.85**	**0.86**
	AUC	0.83	0.82	0.76	0.72	0.62		0.79	0.65	0.61	0.50	0.52
VIX	Psd.R^2	0.09	0.01	0.00	0.01	0.04	Empl.	0.26	0.02	0.00	0.00	0.01
	QPS	0.12	0.13	0.13	0.12	0.11		0.09	0.13	0.13	0.12	0.11
	ACC	0.85	0.83	0.84	**0.85**	**0.86**		0.86	0.83	0.84	**0.85**	**0.86**
	AUC	0.77	0.64	0.46	0.53	0.64		0.86	0.65	0.48	0.53	0.57
S&P 500	Psd.R^2	0.11	0.05	0.00	0.00	0.00	Init. claims	0.33	0.06	0.01	0.00	0.00
	QPS	0.12	0.13	0.13	0.12	0.11		**0.08**	0.13	0.13	0.12	0.11
	ACC	0.84	0.82	0.84	**0.85**	**0.86**		0.89	0.82	0.84	**0.85**	**0.86**
	AUC	0.74	0.71	0.58	0.54	0.55		0.88	0.74	0.65	0.55	0.54
M2	Psd.R^2	0.01	0.00	0.01	0.00	0.00	Unem. rate	**0.35**	0.03	0.00	0.00	0.00
	QPS	0.13	0.14	0.13	0.12	0.11		**0.08**	0.13	0.13	0.12	0.11
	ACC	0.84	0.83	0.84	**0.85**	**0.86**		**0.90**	0.82	0.84	**0.85**	**0.86**
	AUC	0.57	0.57	0.60	0.52	0.51		**0.91**	0.71	0.57	0.51	0.54
Real M2	Psd.R^2	0.02	0.04	0.02	0.04	0.02						
	QPS	0.13	0.13	0.13	0.12	0.11						
	ACC	0.84	0.83	0.84	**0.85**	**0.86**						
	AUC	0.62	0.65	0.61	0.64	0.61						

stock price index and the value of the dollar are also potential covariates. The VIX is included in the model search in recognition that financial volatility may presage a decline in real economic activity.[6] All data span the period January 1973 to December 2013.

Forecasts are evaluated using a variety of metrics. The first is the quadratic probability score (Brier, 1950), which is analogous to squared-error loss for problems with limited dependent variables. The QPS is expressed as

$$\text{QPS} = \frac{1}{T}\sum_{t}(\hat{y}_t - y_t)^2 \tag{9}$$

[6] VIX is taken from CBOE and is available from 1990 to the present. Prior to 1990, VIX is proxied by the within-month standard deviation of the daily return to the S&P 500 index, normalized to the same mean and variance as the VIX for the period when they overlap (1990–2012).

Table III. In-sample forecast accuracy of recession forecasts

		Forecast horizon				
		$h = 0$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
Equally weighted	QPS	0.098	0.120	0.117	0.111	0.109
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
	ACC	0.849	0.826	0.838	0.851	0.859
		(0.03)	(0.05)	(0.04)	(0.04)	(0.04)
	AUC	0.947	0.910	0.862	0.871	0.812
		(0.01)	(0.03)	(0.05)	(0.04)	(0.06)
Bayesian model average	QPS	0.033	0.081	0.090	0.085	0.086
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	ACC	0.951	0.873	0.855	0.857	0.881
		(0.01)	(0.04)	(0.04)	(0.04)	(0.04)
	AUC	0.987	0.923	0.890	0.904	0.846
		(0.01)	(0.03)	(0.04)	(0.04)	(0.05)
Linear boosted	QPS	0.041	0.079	0.088	0.085	0.088
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	ACC	0.941	0.863	0.859	0.857	0.861
		(0.02)	(0.04)	(0.04)	(0.04)	(0.04)
	AUC	0.985	0.934	0.900	0.903	0.846
		(0.01)	(0.02)	(0.04)	(0.04)	(0.06)
Nonlinear boosted	QPS	0.030	0.054	0.062	0.065	0.072
		(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
	ACC	0.959	0.906	0.881	0.902	0.898
		(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
	AUC	0.990	0.970	0.953	0.942	0.915
		(0.01)	(0.01)	(0.02)	(0.03)	(0.05)

Note: Block-bootstrapped standard errors in parentheses. See text for details.

where T is the number of observations and $(\hat{y}_t - y_t)$ denotes the forecast error of the probabilistic forecast to the observation y_t (coded as a 0/1). While intuitive, one shortcoming of the QPS is that it does not evaluate the classification ability of the model. That is, two models may have very different probability scores yet classify 0/1 outcomes in the same way (Hand and Vinciotti, 2003).

The second statistic used to evaluate forecasts is *accuracy*. The accuracy (ACC) is simply the fraction of time that a prediction correctly identifies the state of the business cycle in the sample:

$$\mathrm{ACC} = \frac{1}{T}\sum_t [I(\hat{y}_t \geq c)y_t + (1 - I(\hat{y}_t \geq c))(1 - y_t)] \tag{10}$$

where $\hat{y}$ and y are as before, and c is a threshold that must be specified to turn the probabilistic forecast $\hat{y}_t$ into a binomial forecast. The natural threshold of 50% is used.

Finally, forecasts are evaluated with the area under the receiver operating characteristics (ROC) curve. Whereas accuracy requires a specific threshold to evaluate a probabilistic forecast, the ROC curve describes all possible true positive (TP) and false positive (FP) rates available from a classifier. As the threshold c is varied from 0 to 1, a curve is traced out in $\{\mathrm{TP}(c), \mathrm{FP}(c)\}$ space that describes the classification ability of the model. The area underneath this curve (AUC) is a well-known summary statistic that describes the classification ability of a given model.[7] The statistic has a lower bound of 0.5 and an upper bound of 1, where a higher value indicates superior classification ability. The ROC curve describes the trade-offs between true positive and false negatives produced by a model without imposing a loss function over that trade-off. The AUC is also independent of the incidence of the event being forecast—an important feature when forecasting rare events.

Inference is conducted using a block bootstrap with a fixed window size of 12 months. Experiments with longer window sizes (not reported) showed that lengthening the window size has little influence on the reported results, indicating that 12 months sufficiently captures any autocorrelation present in the data.

[7] See Pepe (2003) for a complete introduction.

IN-SAMPLE RESULTS

To motivate the problem of model selection, Table II presents in-sample measures of model fit for univariate logistic models estimated at forecast horizons up to 24 months. For each covariate and forecast horizon, the table shows four summary statistics: the pseudo-R^2 of Estrella and Mishkin (1998), the QPS, accuracy and the AUC. The best-performing covariate according to each statistic is shown in bold.

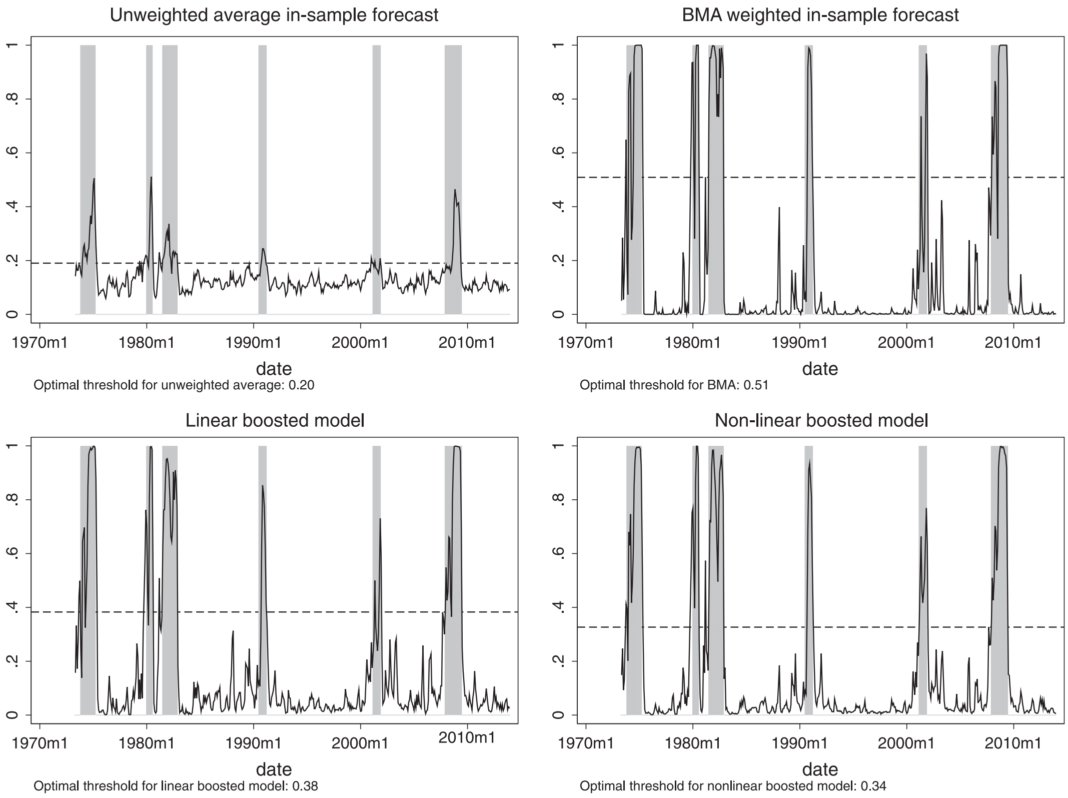


Figure 1. In-sample recession probabilities forecast at horizon of zero months. The top panels contain forecasts from model combination schemes, equally weighted on left and Bayesian model averaged at right. The two bottom panels contain recession probabilities from linear (left) and nonlinear (right) boosted forecast models. Grey shading indicates NBER-defined recession dates and dashed line denotes optimal threshold. See text for details

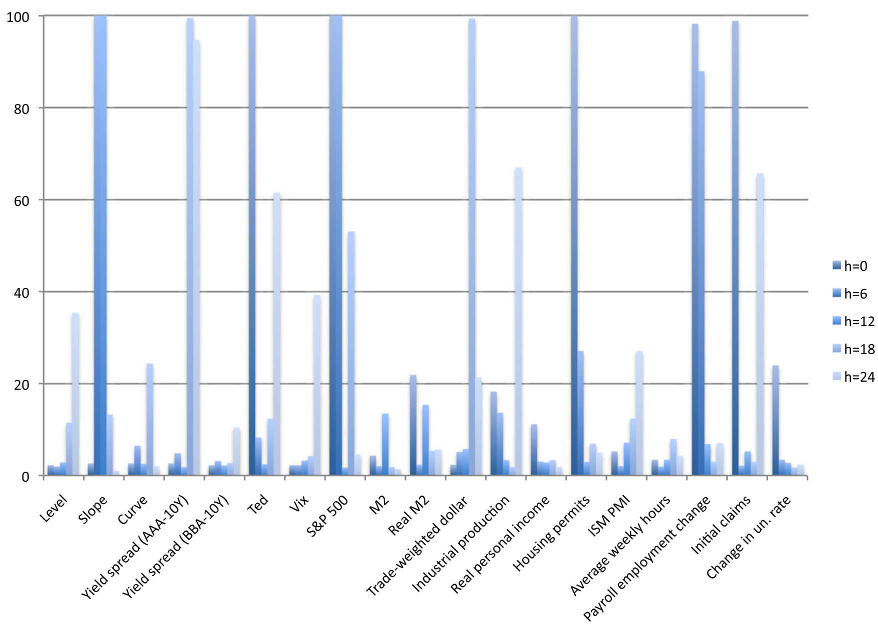


Figure 2. Covariates selected by Bayesian model averaging by forecast horizon (in-sample). The figure displays the posterior inclusion probability (PIP) for the BMA exercise. See text for details

The indicators show large differences in forecast ability. Variables that describe real economic activity identify recessions at short horizons—an observation that is unsurprising since the NBER recession dates themselves are based on the behavior of these variables. Industrial production, monthly employment gains, initial claims and the change in unemployment rate are all useful indicators of the state of the economy at short horizons. These variables have high pseudo-R^2s and AUC statistics near 0.90 when nowcasting. Model fit declines as the forecast horizon grows for these indicators, and many forecast recessions as poorly as a coin toss 12 months ahead. In contrast, many of the financial indicators exhibit forward-looking behavior but cannot forecast at short horizons.

It is worth noting how each evaluation metric describes in-sample forecast ability, especially at very long horizons. Many of the models do not provide clear recession signals at long horizons, in the sense that the implied probabilities do not cross the 50% threshold. However, because the incidence of recessions in the USA is relatively low—16% of months in the sample—these models nevertheless obtain a high accuracy rate. Yet the models have a very low (or even zero) true positive rate, and are of little use when forecasting downturns. QPS is similarly sensitive to the incidence of recessions. In contrast, the AUC is independent of the incidence of recessions when evaluating the forecast ability of the models.

Table III evaluates the forecasts produced by the methods described under ‘Empirical setup’ (above); the implied recession probabilities are shown in Figure 1. Equally weighted forecasts classify recessions quite well, with AUC statistics at each horizon that are similar to or higher than the best-performing univariate model. However, as seen in

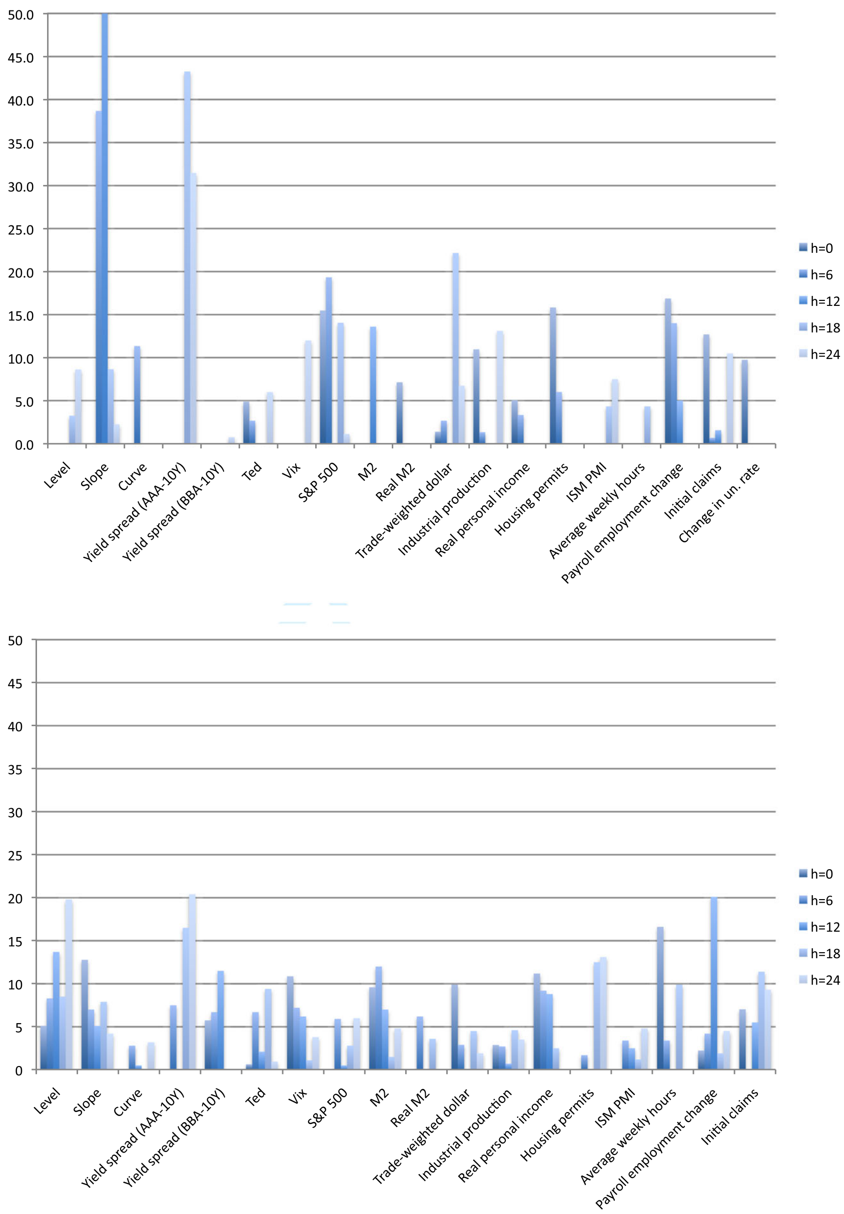


Figure 3. Covariates selected by boosted models by forecast horizon (in-sample). The panels display the fraction of iterations for which a particular covariate was selected by the boosting algorithm, linear (top) and nonlinear (bottom). See text for details

Table IV. Business cycle turning points, 1973–2012

NBER	Unweighted average	BMA	Linear boost	Nonlinear boost
Peak dates				
1973:11	2	7	2	2
1980:1	−3	3	3	−3
1981:7	0	0	0	0
1990:7	2	2	2	−2
2001:3	—	—	—	0
2007:12	7	8	0	1
Trough dates				
1975:3	1	2	2	−2
1980:7	1	0	1	0
1982:11	0	0	0	0
1991:3	−1	0	0	1
2001:11	—	—	—	1
2009:6	0	0	0	0

Note: Value shown is the model-implied peak/trough calculated using the optimal threshold and a rule that each phase of the business cycle lasts more than 3 months. The value listed is the model-implied peak/trough date relative to the NBER-defined peak/trough date. For example, a positive value of 3 indicates that the model dated the peak or trough 3 months later than the NBER (e.g. May instead of February). '—' indicates that the model did not generate a business cycle phase lasting 3 months that corresponds to the NBER recession. See text for details.

panel (a) of Figure 1, the equally weighted forecast hovers around the unconditional probability of recession. That the implied recession probability is indeed higher during recessions produces a high AUC statistic. The model never gives a very strong signal of recession—the maximum probability of recession is only about 0.50.

The second, third and final panels of Table III and Figure 1 show that the in-sample performance of the BMA and boosted models are impressive. Contemporaneously, each has an AUC that indicates near-perfect classification ability. As the forecast horizon lengthens, the BMA and linear boosted models exhibit very similar forecast ability, with performance similar to or better than the best-performing univariate model at each horizon. The performance of the nonlinear models, at least in-sample, is superb. At each forecast horizon, the nonlinear specification outperforms the best univariate model and each of the other model selection schemes. The difference between forecasting models is most notable at horizons of 12 months and higher.

Figures 2 and 3 provide insight into the model selection. Figure 2 shows posterior inclusion probabilities (PIPs) for each forecast horizon.[8] BMA models rely only on a handful of covariates for each forecast horizon, but the covariates vary by forecast horizon. Contemporaneously, employment data, TED spread, S&P 500 and housing data have PIPs above 95%. In contrast, models of recession 6 and 12 months into the future rely on the slope of the yield curve. The model that forecasts 12 months ahead assigns a PIP to the slope of the yield curve of 100%, while no other indicator receives a PIP higher than 10%. The models forecasting at 18 and 24 months are dominated by corporate yield spreads.

The models selected by the boosting algorithms, shown in Figure 3, are similar to those from BMA. Figure 3 shows the fraction of iterations that the boosting algorithm selected a particular covariate.[9] The boosted models include more covariates than BMA, but the covariates selected for each horizon are qualitatively similar. At short horizons, the forecast model relies on measures of real activity. In the medium term, the model conditions on the slope of the yield curve. Corporate yield spreads are included in models that forecast at long horizons. Nonlinear models are typically similar to their linear counterparts, but do have some covariates that were excluded from the linear case. For example, VIX, which was rarely included in the linear model, enters into the nonlinear model at all forecast horizons, receiving a heavy weight contemporaneously.

In-sample recession probabilities can be combined with a threshold value to produce a chronology of business cycle turning points for the USA economy. For each series displayed in Figure 1, an optimal threshold is produced by assuming symmetric utility/disutility from true/false positives. Under this symmetry assumption, the utility of a classifier is the difference between true and false positive rates (Berge and Jordà, 2011). The line shown in the figure is the threshold that maximizes this difference for each model. Peaks and troughs are the first and final month for which the recession probability is equal to or greater than the threshold, with the additional restriction that each phase

[8] The PIP is a weighted average of the posterior probabilities for each of the 2^K models that includes covariate j; that is, $\text{PIP}(\beta_j) = \Pr(\beta_j \neq 0) = \sum_{M_i:\beta_j \in M_i} p(M_i|D)$.

[9] The plot is of ψ_k^h for each model horizon h and covariate k, where $\psi_k^h = \frac{1}{I}\sum_{i=1}^{I} I(k = \kappa)$ and κ denotes the covariate with minimum mean squared error at iteration i. Note that the selection frequencies from the boosting algorithm sum to 100% across indicators.

of the business cycle lasts more than 3 months. Table IV displays the difference between those peak and trough dates relative to the NBER recession dates.

The model-implied recession dates align with the NBER dates quite well. While there are no false positives in the sample, the 2001 recession remains difficult to recognize. During that episode, each model produces a predicted probability that crosses its relevant threshold, but only the nonlinear boosted model produces a predicted probability that stays above the threshold for more than three consecutive months. The unweighted average performs least well of the chronologies. The BMA model does not clearly identify the start of the 1973 and 2007 recessions but otherwise performs similarly to the boosted models.

The key lesson from the analysis in this section is that the information carried by economic indicators varies widely by forecast horizon. When forecasting the contemporaneous probability of recession, many of the established indicators of recession are found to carry little information and practitioners should focus on indicators of real economic activity. Conversely, the indicators that forecast best at short horizons often cannot forecast a recession 12 months into the future accurately. Although model averaging can alleviate concerns of model misspecification, the average still depends critically on the underlying indicators and the weights assigned to each covariate.

OUT-OF-SAMPLE PERFORMANCE

A risk of model search is that the models may overfit the data, limiting out-of-sample forecast ability. This section considers two out-of-sample forecast experiments. The first exercise uses recent-vintage data, producing forecasts using a rolling window estimator for the period May 1985 to December 2013. This experiment focuses on the stability and average forecast ability of the forecast models. In addition, since many of the variables that describe real activity are subject to revision, a second experiment considers the real-time performance of the forecast methods surrounding the 2001 and 2007 recessions. Data for the real-time forecasting exercise is from the ALFRED website, maintained by the Federal Reserve Bank of St Louis. Recession models that rely on NBER recession dates are confounded by the fact that NBER announcements are typically delayed 12–18 months after the turning point. The second experiment uses real-time NBER dates, assuming that the dependent variable does not change until the NBER has made an official announcement declaring each turning point. This approach is conservative since it may be clear to practitioners that a

Table V. Out-of-sample rolling window forecast performance

		Forecast horizon				
		$h = 0$	$h = 6$	$h = 12$	$h = 18$	$h = 24$
	N	344	344	344	344	344
Equally weighted	QPS	0.072	0.088	0.090	0.088	0.091
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
	ACC	0.911	0.893	0.890	0.890	0.890
		(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
	AUC	0.844	0.784	0.775	0.787	0.751
		(0.08)	(0.07)	(0.07)	(0.09)	(0.08)
BMA	QPS	0.041	0.085	0.082	0.068	0.081
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	ACC	0.950	0.890	0.896	0.911	0.896
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
	AUC	0.958	0.836	0.884	0.905	0.808
		(0.02)	(0.06)	(0.03)	(0.04)	(0.07)
Linear boosted	QPS	0.033	0.080	0.078	0.076	0.087
		(0.01)	(0.02)	(0.03)	(0.02)	(0.03)
	ACC	0.958	0.887	0.890	0.890	0.890
		(0.02)	(0.04)	(0.04)	(0.04)	(0.04)
	AUC	0.962	0.908	0.868	0.887	0.765
		(0.02)	(0.03)	(0.06)	(0.06)	(0.08)
Nonlinear boosted	QPS	0.027	0.070	0.069	0.068	0.075
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
	ACC	0.964	0.896	0.896	0.896	0.902
		(0.01)	(0.03)	(0.04)	(0.04)	(0.04)
	AUC	0.966	0.923	0.943	0.926	0.864
		(0.02)	(0.03)	(0.02)	(0.04)	(0.05)

Note: Initial forecast is May 1985 and forecast window continues to December 2013. Block-bootstrapped standard errors in parentheses.

turning point has occurred well before the NBER announces it. As before, forecasts are evaluated using the quadratic probability score, accuracy and area under the ROC curve.

Out-of-sample forecast performance is diminished relative to, but consistent with, the in-sample results. As shown in Table V, each forecast method produces useful forecasts out-of-sample. With the caveat that these forecasts use recent-vintage data, Table V shows that short horizon forecasts, on average, perform quite well. Unsurprisingly, forecast performance deteriorates as the forecast horizon lengthens. However, the performance of the models forecasting at medium and longer horizons—especially the models forecasting 12 months ahead—are quite good, with AUCs near 0.90. These models rely on financial data so that data revisions are less serious an issue forecasting at this horizon.

Figures 4 and 5 show the variables deemed to be important predictors of recession, and explores the stability of those models. Figure 4 shows the PIPs for the five variables with the largest average PIP in the out-of-sample exercise at forecast horizons zero (top) and 12 (bottom). PIPs for variables not shown are considerably less volatile, and are generally well below 10%. The top panel suggests instability in the nowcasting model following the 1991, 2001 and 2007 recessions. During each recession, the model selects a new covariate or drops another. For example, the model begins to include the Ted spread following the 2007 recession, which had been only sporadically included previously. The posterior probabilities of the model forecasting 12 months ahead also appear to undergo regime shifts as the model moves through the sample window. Throughout the 2000s, the BMA forecast relies primarily on the slope of the yield curve. Although the financial crisis adds additional covariates into the model, they largely disappear by the end of the sample.

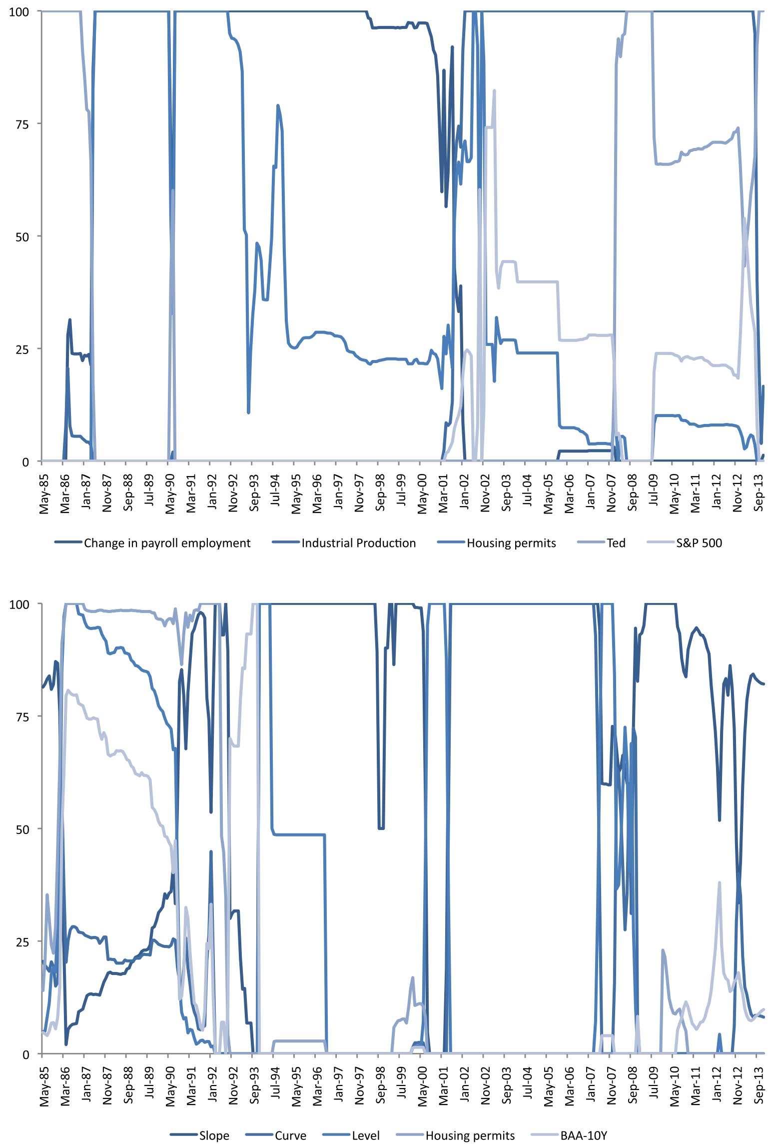


Figure 4. Posterior inclusion probabilities for BMA models used to produce out-of-sample forecasts. The figure shows the PIP for the five indicators with the highest average PIP in the rolling window exercise. The top panel is for the model producing nowcasts, while the bottom panel shows the model forecasting 12 months ahead. See text for details

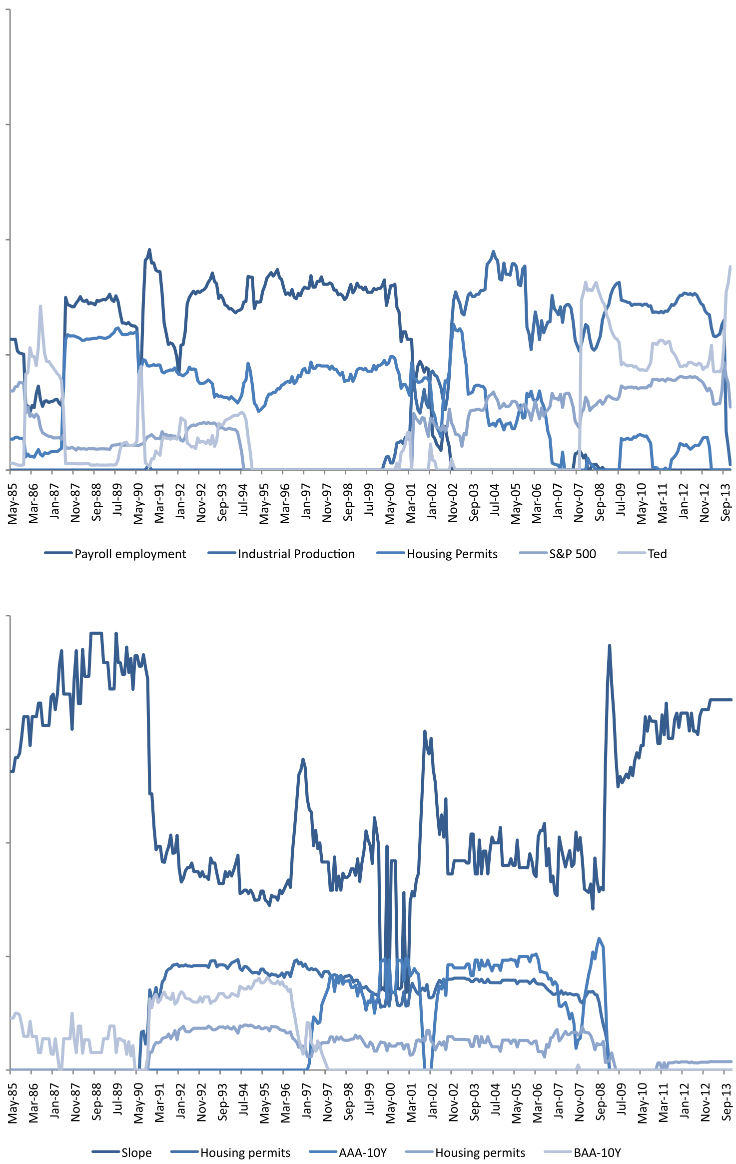


Figure 5. Out-of-sample selection frequency of covariates for linear boosted model. The figure shows the average selection frequency for the five indicators with the highest average selection frequency in the rolling window exercise. The top panel is for the model producing nowcasts, while the bottom panel shows the model forecasting 12 months ahead. See text for details

The boosted models also appear to capture regime changes. Figure 5 displays the frequency of times a given covariate is selected by the linear boosting algorithm forecasting at 0 and 12 months. As in Figure 4, the figure shows the five variables with the highest average probability of inclusion. The out-of-sample models qualitatively resemble those from the in-sample analysis. At short horizons, the boosting model selects variables that describe real activity, primarily the change in payroll employment, but also industrial production and housing starts. As was the case with the BMA model, the Ted spread enters the model after 2007. At the 12-month horizon, the model relies primarily on the slope of the yield curve and other yield spreads, but also a measure of the housing market.

Finally, Figures 6 and 7 show real-time, out-out-sample forecasts surrounding the two most recent recessions. The figures show probabilistic recession forecasts at horizons of 0 (top panel) and 12 months (bottom panel)—i.e. $P(\text{NBER}_t = 1|D_{t-h-1})$—in 3-month increments beginning 1-year prior to the NBER-defined recession. For the sake of comparison, real-time forecasts from alternative forecasting models are included in Figures 6 and 7. Forecasts from the model of Chauvet and Piger (2008) are shown as an alternative nowcast. The Chauvet–Piger model is state-of-the-art, a factor Markov-switching model extracted from four variables that describe real activity. However, because the Chauvet and Piger (2008) model is intended to identify, not forecast, recession events, the 12-month-ahead probabilities are compared to the best-performing univariate model, the model using the slope of the yield curve, also estimated in real time. Indeed, the literature recognized the slope of the yield curve may have signaled a downturn prior to the 2007 recession (Wright, 2006).

No model provides an obvious nowcast of the 2001 recession in real time. The Chauvet–Piger model produces recession probabilities that most closely align with the NBER dates—the probabilities exceed 30% beginning in

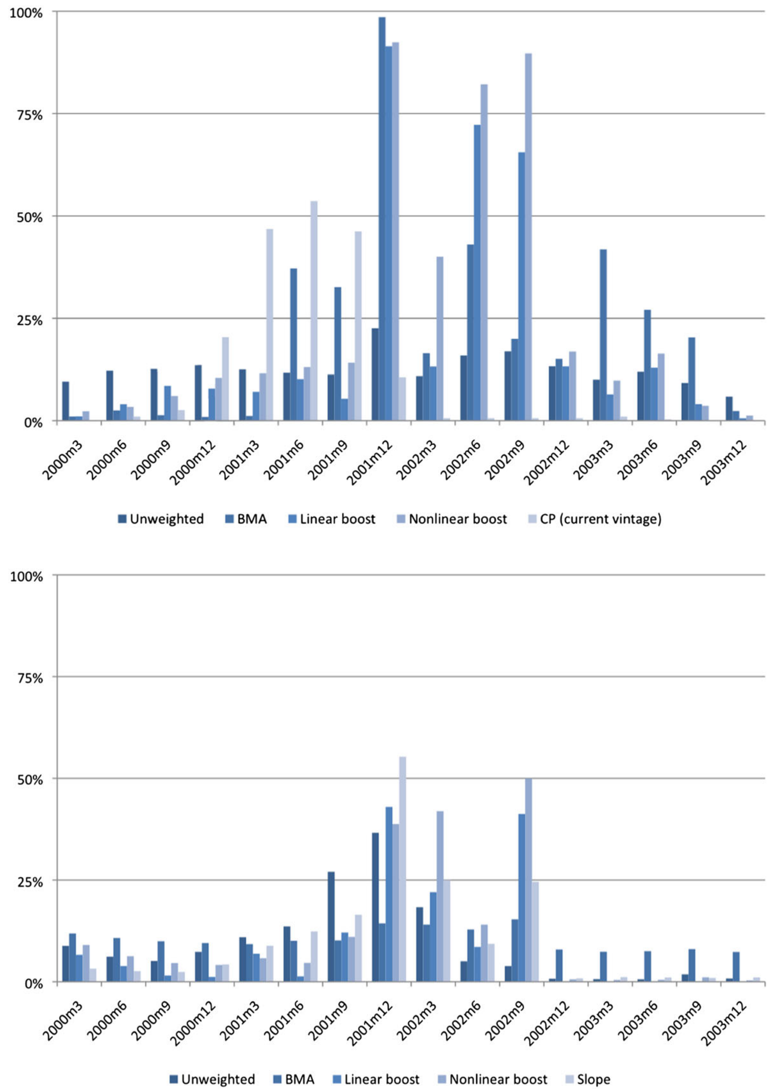


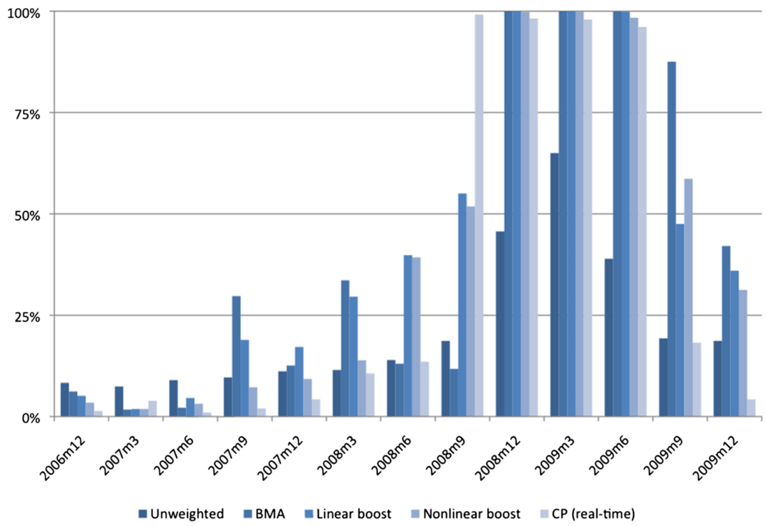
Figure 6. Real-time out-of-sample recession probabilities surrounding 2001 recession, i.e. Pr(NBER$_t$ $= 1|D_{t-h-1}$). Nowcasts ($h = 0$) are shown in the top figure. Included for comparison are Chauvet and Piger (2008) recession probabilities. The bottom figure shows the real-time 1-year-ahead recession probabilities. Real-time probabilities from a univariate logit model using the slope of the yield curve are also shown. Official NBER recession dates are March 2001 to November 2001

January 2001 and through November 2001. Reflecting that the 2001 recession was relatively mild, the predicted probability of recession is never very strong, and only barely exceeds 50% at its peak in June 2001. Of the other models considered, the BMA and two boosted models provide somewhat stronger signals of recession, but not until December 2001, 1 month after the NBER-defined trough in activity.

Turning to the 12-month-ahead forecast surrounding the 2001 recession, again no model provided a strong signal of recession. The strongest signals occurred around December 2000 but, as above, the signals barely exceed 50%. The lack of a strong signal from each forecast models in 2001 is consistent with the previous literature, especially for those models forecasting in the medium term and using the yield curve (Chauvet and Potter, 2005).

The results for identifying the 2007 recession in real time are also mixed. Figure 7 shows real-time recession probabilities for each of the four models considered, as well as the Chauvet–Piger recession probabilities. Each model identifies a clear recession event, but with a substantial delay relative to the NBER-defined peak of economic activity (December 2007). The recession probabilities remain elevated throughout 2008 and some well into 2009. These results are consistent with the substantial downward revisions that occurred to many economic variables during 2008. Likely reflecting the sluggish labor market recovery that followed this recession, each of the models remain elevated somewhat longer following the NBER-defined trough of activity (June 2009), although each recedes below 50% by December 2009.

The bottom panel shows that no forecasting model produced a strong signal ahead of the recession. The recession probabilities produced by models forecasting 12 months ahead rose beginning in early 2007, but the signals are not acute and do not exceed 40%. Since the models rely largely on the slope of the yield curve, it is worth noting that the yield curve did invert for a brief period in late 2006 and early 2007, but achieved a minimum of only −40 bps,



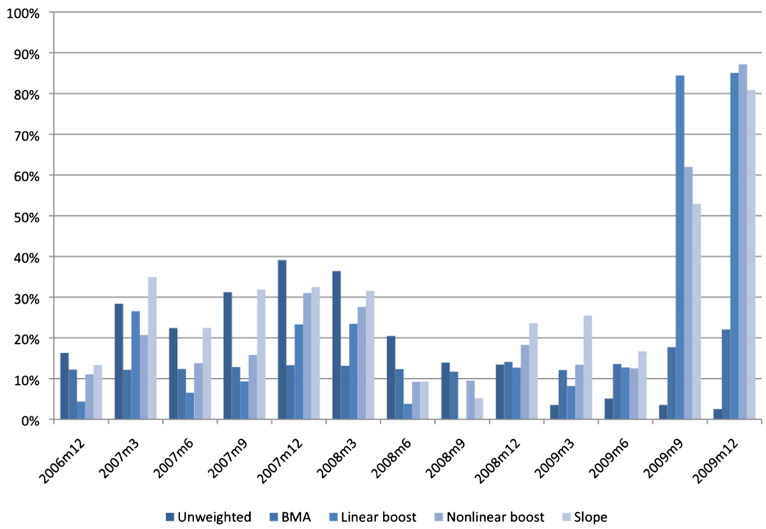


Figure 7. Real-time out-of-sample recession probabilities surrounding 2007 recession, i.e. Pr(NBER$_t$ $= 1|D_{t-h-1}$). Nowcasts ($h = 0$) are shown in the top figure. Included for comparison are Chauvet and Piger (2008) recession probabilities. The bottom figure shows the real-time 1-year-ahead recession probabilities. Real-time probabilities from a univariate logit model using the slope of the yield curve are also shown. Official NBER recession dates are December 2007 to June 2009

which, relative to prior recessions, is not a strong warning signal. Similarly, while housing permits were negative throughout the fall of 2006, they turned briefly positive in early 2007 before the bottom fell out in late 2007 and 2008. The model selection schemes forecasting 12 months provided signals of recession that are similar to the simple yield curve model prior to the 2007 recession. This is not surprising since, as can be seen in Figures 4 and 5, the models relied on the slope of the yield curve in the run-up to this recession.

CONCLUSION

This paper evaluates the information content of many commonly cited economic indicators to establish empirical regularities about the variables that lead business cycle turning points. Two methodologies—model averaging and model selection—are used to produce horizon-specific forecasting models. Covariates included in each forecast model are unique to each forecast horizon. In contrast to large portions of the forecasting literature, the analysis showed that a simple model average may not be appropriate for recession forecasts. Since many commonly followed indicators are valuable at only particular forecast horizons, a simple model average of many indicators dilutes useful information. Empirically driven model selection algorithms such as Bayesian model averaging and boosting are more successful, but cannot overcome the fact that the best forecasting models change over time.

The analysis also serves as a point of comparison for methods of model selection. Equally weighted and Bayesian model averaging are well established in the forecasting literature, but boosting is not as well known. The variant of the algorithm applied here is a stage-wise additive logistic model and should look very familiar to economists. When applied in a real-time forecasting environment, the results from boosting are similar to those of BMA. At short horizons, both BMA and boosting have modest success identifying turning points, and while they offer an alternative to factor regime-switching models they do not outperform the Chauvet–Piger model when nowcasting the 2001 and 2007 turning points. They do show modest improvement over forecasts produced by a yield curve model when forecasting the two most recent recessions 12 months ahead. Boosting has many appealing properties: it is computationally efficient, intuitive and easily applied to very large datasets where BMA may be infeasible.

Overall, the results indicate that there is no sufficient summary statistic for forecasting business cycle turning points. Indeed, this is the approach taken by the NBER when identifying turning points; the committee considers a variety of economic indicators when making pronouncements about the state of the economy. The power of the yield curve as a predictor of future economic activity largely endures, in the sense that models selected to forecast recession 1 year ahead rely heavily on this indicator. At the same time, no model gave strong warning signals ahead of the 2001 and 2007 recessions. This observation should give pause to those forecasting future recessions.

ACKNOWLEDGEMENTS

The views herein do not necessarily reflect the views of the Federal Reserve System. Seminar participants at the 2012 ISF, Federal Reserve Bank of Kansas City and Federal Reserve Board as well as two anonymous referees provided very useful feedback. Of course, any mistakes are my own.

REFERENCES

Aruoba SB, Diebold FX, Scotti C. 2009. Real-time measurement of business conditions. *Journal of Business and Economic Statistics* **27**(4): 417–427.
Bai J, Ng S. 2009. Boosting diffusion indices. *Journal of Applied Econometrics* **24**(4): 607–629.
Bates J, Granger C. 1969. The combination of forecasts. *Operations Research Quarterly* **20**: 451–468.
Berge TJ. 2014. Forecasting disconnected exchange rates. *Journal of Applied Econometrics* **29**(5): 713–735.
Berge TJ, Jordà O. 2011. Evaluating the classification of economic activity. *American Economic Journal: Macroeconomics* **3**: 246–277.
Brier GW. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* **75**: 1–3.
Brock WA, Durlauf SN, West KD. 2007. Model uncertainty and policy evaluation: Some theory and empirics. *Journal of Econometrics* **136**(2): 629–664.
Buhlmann P, Hothorn T. 2007. Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science* **22**(4): 477–505.
Buhlmann P, Yu B. 2003. Boosting with the l2 loss: Regression and classification. *Journal of the American Statistical Association* **98**: 324–340.
Burns AF, Mitchell WC. 1946. *Measuring Business Cycles*. National Bureau of Economic Research: New York.
Chauvet M. 1998. An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review* **39**(4): 969–996.
Chauvet M, Hamilton JD. 2006. Dating business cycle turning points. In *Nonlinear Time Series Analysis of Business Cycles*, vol. 276, Milas C, Rothman PA, van Dijk D, Wildasin DE (eds), Contributions to Economic Analysis. Emerald Group: Bingley, UK; 1–54.
Chauvet M, Piger J. 2008. A comparison of the real-time performance of business cycle dating methods. *Journal of Business and Economic Statistics* **26**: 42–49.
Chauvet M, Potter S. 2002. Predicting a recession: evidence from the yield curve in the presence of structural breaks. *Economics Letters* **77**(2): 245–253.
Chauvet M, Potter S. 2005. Forecasting recessions using the yield curve. *Journal of Forecasting* **24**(2): 77–103.
Chauvet M, Senyuz Z. 2012. A dynamic factor model of the yield curve as a predictor of the economy. *Technical Report 2012-32*. Federal Reserve Board.
Chen Z, Iqbal A, Lai H. 2011. Forecasting the probability of US recessions: A probit and dynamic factor modelling approach. *Canadian Journal of Economics* **44**(2): 651–672.
Dueker M. 2005. Dynamic forecasts of qualitative variables: A qual var model of U.S. recessions. *Journal of Business and Economic Statistics* **23**: 96–104.
Eilers PH, Marx BD. 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* **11**(2): 89–121.
Elliott G, Lieli RP. 2013. Predicting binary outcomes. *Journal of Econometrics* **174**(1): 15–25.
Estrella A, Mishkin FS. 1998. Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics* **80**(1): 45–61.
Fossati S. 2011. Dating U.S. business cycles with macro factors. Working papers 2011-5. Department of Economics, University of Alberta.
Friedman J. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **29**(5): 1189–1232.
Giusto A, Piger J. 2013. Nowcasting U.S. business cycle turning points with vector quantization. Working paper. University of Oregon.
Gurkaynak RS, Sack B, Wright JH. 2007. The U.S. treasury yield curve: 1961 to the present. *Journal of Monetary Economics* **54**(8): 2291–2304.
Hamilton JD. 2011. Calling recessions in real time. *International Journal of Forecasting* **27**(4): 1006–1026.
Hand DJ, Vinciotti V. 2003. Local versus global models for classification problems: Fitting models where it matters. *American Statistician* **57**: 124–131.
Hendry DF, Clements MP. 2004. Pooling of forecasts. *Econometrics Journal* **7**: 1–31.
Kauppi H, Saikkonen P. 2008. Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics* **90**(4): 777–791.

Khandani A, Kim AJ, Lo AW. 2010. Consumer credit risk models via machine-learning algorithms. *Technical report*. MIT Working Paper Series.
Marcellino M, Stock JH, Watson MW. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499–526.
Morley J, Piger J. 2012. The asymmetric business cycle. *Review of Economics and Statistics* **94**(1): 208–221.
Ng S. 2014. Boosting recessions. *Canadian Journal of Economics* **47**(1): 1–34.
Ng S, Wright JH. 2013. Facts and challenges from the great recession for forecasting and macroeconomic modeling. NBER Working Paper 19469. National Bureau of Economic Research.
Oreopoulos P, Page M, Stevens AH. 2006. The intergenerational effects of worker displacement. *Journal of Labor Economics* **24**: 455–483.
Owyang MT, Piger J, Wall HJ. 2013. Forecasting national recessions using state level data. Working paper series 2012-013B. Federal Reserve Bank of St Louis.
Pepe MS. 2003. *The Statistical Evaluation of Medical Test for Classification and Prediction*. Oxford University Press: Oxford.
Raftery AE. 1995. Bayesian model selection in social research. *Sociological Methodology* **25**: 111–163.
Rudebusch GD, Williams JC. 2009. Forecasting recessions: The puzzle of the enduring power of the yield curve. *Journal of Business and Economic Statistics* **27**(4): 492–503.
Ruhm CJ. 1991. Are workers permanently scarred by job displacements? *American Economic Review* **81**(1): 319–324.
Ruhm CJ. 2000. Are recessions good for your health? *Quarterly Journal of Economics* **115**(2): 617–650.
Sala-i-Martin X, Doppelhofer G, Miller RI. 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review* **94**(4): 813–835.
Stevens AH, Schaller J. 2011. Short-run effects of parental job loss on children's academic achievement. *Economics of Education Review* **30**(2): 289–299.
Stevens AH, Miller D, Page M, Filipski M. 2011. The best of times, the worst of times: Understanding pro-cyclical mortality. NBER Working Paper 17657.
Stock J, Watson M. 1989. New indexes of coincedent and leading economic indicators. *NBER Chapters, in: NBER Macroeconomics Annual 1989* **4**: 351–409. National Bureau of Economic Research.
Stock JH, Watson MW. 1999. Forecasting inflation. *Journal of Monetary Economics* **44**(2): 293–335.
Stock JH, Watson M. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**(6): 405–430.
Sullivan DG, von Wachter T. 2009. Job displacement and mortality: An analysis using administrative data. *Quarterly Journal of Economics* **124**(3): 1265–1306.
Timmermann A. 2006. Forecast combinations. In *Handbook of Economic Forecasting*, Elliott G, Granger C, Timmermann A (eds). North-Holland: Amsterdam; 135–196.
Wright JH. 2006. The yield curve and predicting recessions. *Technical report*. Board of Governors of the Federal Reserve Board.

Author's biography:
Travis J. Berge is an Economist in the division of Research and Statistics at the Federal Reserve Board.

Author's address:
Travis J. Berge, Board of Governors of the Federal Reserve System, 20th St and C St NW, Washington, DC 20431, USA.

APPENDIX: DATA SOURCES

Table AI. Data sources

Variable	Available dates	Source
	Interest rates	
Effective Fed funds rate	1954:7–2013:12	FRB H.15 release
3-month eurodollar deposit rate	1971:1–2013:12	FRB H.15 release
3-month Treasury yield (secondary market)	1934:1–2013:12	Gurkaynak *et al.* (2007)
1-year Treasury yield (constant maturity)	1953:4–2013:12	Gurkaynak *et al.* (2007)
10-year Treasury yield (constant maturity)	1953:4–2013:12	Gurkaynak *et al.* (2007)
Moody's BAA corporate bond yield	1947:1–2013:12	FRB H.15 release
	Other financial variables	
S&P 500 Index	1947:1–2013:12	Standard & Poors
M2 (seasonally adj.)	1959:1–2013:12	FRB H.6 release
Trade-weighted US dollar: major currencies	1973:1–2013:12	FRB H.10 release
	Macroeconomic indicators	
CPI index (seasonally adj.)	1947:1–2013:12	US Dept of Labor: BLS
Industrial production (seasonally adj.)	1947:1–2013:12	FRB G.17 release
Real personal income (seasonally adj.)	1959:1–2013:12	US Dept of Commerce
ISM manufacturing PMI index	1948:1–2013:12	Institute for Supply Management
Housing permits	1960:1–2013:12	US Census Bureau
Average weekly hours: manufacturing	1947:1–2013:12	US Dept of Labor: BLS
All employees (total nonfarm)	1939:1–2013:12	US Dept of Labor: BLS
Initial claims for unemployment insurance	1967:1–2013:12	US Dept of Labor: BLS
Civilian unemployment rate	1948:1–2013:12	US Dept of Labor: BLS