Nowcasting Recessions with Machine Learning: New Tools for Predicting the Business Cycle

Jeremy Garbellano

June 3, 2016

A THESIS

Presented to the Department of Economics in partial fulfillment of the requirements for the degree of Bachelor of Science
Supervised by Professors Jeremy Piger and Tim Duy

APPROVED: ______________________________
Professor JEREMY PIGER

I Introduction

A fundamental attribute of capitalist economies is their propensity to undergo cyclical periods of expansion and contraction in output. The modern study of economic cycles was initiated in 1819 by Jean Charles Léonard de Sismondi, who posited that business activity did not maintain a consistent equilibrium between supply and demand, but instead alternated between periods of positive and negative growth due to mismatches in production, consumption and labor market dynamics. Around the same period, Thomas Malthus proposed that bouts of economic stagnation ("gluts") were structural to capitalism, and arose from disparities in wealth distribution and consumption preferences among laborers, landowners, and capitalists. Before Sismondi and Malthus, classical economists relied on exogenous forces, e.g., war and famine, to rationalize the presence of cyclical behavior, or more commonly, ignored short-run dynamics altogether.

Today, a prevailing view among economists is that expansions and recessions represent unobserved variables to be estimated according to the data (Fossati 2016). While policy-oriented economists seek to explain and reconcile issues brought on by unexpected shifts in the business cycle, econometric approaches concentrate on identifying economic peaks and troughs, as well as predicting turning points. To this end, business cycle practitioners address four related questions:

1. Given sufficient hindsight, which periods can be classified as expansionary versus recessionary?
2. Based on the most recent economic data, what is the current state of the economy?
3. Which economic indicators are predictive of future regime shifts?
4. What is the extent and time horizon for which one can reliably forecast regime changes?

For much of the 20th century, efforts were focused primarily on the first question, with economists Arthur Burns and Wesley Mitchell standing out as modernizing the field with their 1946 book, *Measuring Business Cycles.* Prior to the book's publication, business cycle dating methodologies often viewed cyclical behavior from a literalist perspective (i.e., as exhibiting a fixed duration and amplitude), and sought to identify such patterns in a small set of variables deemed representative of the broader economy (Zarnowitz 1992). Burns and Mitchell, in contrast, observed cyclical fluctuations as "recurrent but not periodic" and "widely diffused over the economy" - views that not only remain well-accepted today, but also continue to drive new research. To determine chronologies for expansions and recessions, the two economists identified individual turning point dates across a large, diverse set of economic variables, and then analyzed the distribution of these dates to arrive at a reference date signaling the turning point for the broader economy.

The National Bureau of Economic Research (NBER), a private, nonprofit institution founded by Wesley Mitchell in 1920, stands today as the *de facto* authority in classifying periods of expansion and recession. To facilitate this effort, NBER established the Business Cycle Dating Committee (BCDC) in 1978 to issue public statements regarding the start and end of recessionary periods, which are defined as "significant decline[s] in economic activity spread across the economy" that last "from a few months to more than a year." Eschewing the use of a fixed rule to classify recessions, NBER economists instead rely on internal judgment over "the behavior of various measures of broad activity" - namely, gross domestic product, employment, personal income, retail sales and industrial production. Since 1980, the BCDC has issued 10 press releases, with business cycle peak and trough date announcements released between 6 and 21 months following their actual occurrence.

Although NBER's stature as a centralized, nonpartisan authority averts the possibility of having competing sets of expansion and recession dates and ensures conflicting political or financial motives do not influence the task at hand, the institution's implicit cost function stands in contrast to other economic stakeholders. Specifically, because the BCDC is concerned foremost with maintaining accurate business cycle dating chronologies in the face of noisy economic data and multiple *ex post* revisions [1], their inclination towards conservatism increases requirements for confirmatory data in the months following initial detection of a regime shift. Evidently, this presents considerable uncertainty in the public sphere in the interim periods between when turning points occur and when they are officially announced.

Since regime shifts influence the activities of individuals, investors, employers and policymakers, such groups might weigh other considerations above NBER's primary focus on accuracy. A university economist, for example, might value transparency and reproducibility above all else, leading to demand for dating methods that rely on explicit quantitative rules and/or statistical models. Investors, for their part, often place a premium on access to advanced knowledge before it becomes embedded in market consensus. As a result, a portfolio manager might prefer more timely warnings of impending regime changes to guide investment decisions - even if such information is accompanied by occasional false signals. Firms and individuals, in contrast, might place equal weighting on both timeliness and accuracy to inform decisions regarding employment, budgeting and financing, given their relatively higher costs associated with acting on inaccurate information.

As a result of these differing preferences, econometricians have devoted considerable effort to devising methods of classifying economic expansions and recessions independent of NBER's official proclamations. These techniques have generally fallen into three categories - dating, forecasting and nowcasting. Dating methodologies, in following the legacy set by Burns and Mitchell, seek to apply explicit rules and/or quantitative models to economic data in order to classify expansions and recessions in near-complete agreement with NBER's

[1] Q4 2015 US GDP, for instance, was announced to be 0.7% four weeks following the quarter end. By March 2016, the Commerce Department revised this figure to 1.4%, a nearly 100% increase

official dates (see, for example, Harding and Pagan 2003; Chauvet and Piger 2008; Stock and Watson 2010). Accomplishing this feat generally requires waiting for an extended period following the detection of a turning point to minimize or eliminate false signals. Forecasting methodologies estimate when the next regime shift will occur via statistical models and a carefully-selected set of leading and coincident indicators. Empirically, such approaches have generally focused on 3-12 month horizons, as out-of-sample performance tends to deteriorate beyond that horizon (Berge 2013). Nowcasting techniques, an intermediate step between retroactively dating business cycles and forecasting new ones, classify whether the economy is currently in an expansion or recession based on real-time data releases. Their primary *raison d'etre* is to take advantage of the lag between the start and end dates of recessions and their official announcement by NBER in providing an advanced (albeit still reliable) signal of impending regime shifts.

While the question of when the next recession will occur is typically viewed as more interesting than whether the economy is currently facing one, addressing the former presents a number of challenges. First, few variables hold predictive value beyond the 6 month mark, with the slope of the yield curve[2] and corporate credit spreads standing out as primary candidates (see, for example, Stock and Watson 1989; Estrella and Mishkin 1996; Dueker 1997; Rudebusch and Williams 2008; Berge 2013). Second, even with an optimal variable selection, out-of-sample recession probabilities might yet fail to breach the 50% decision threshold[3] commonly used in the literature - a trait seen, among others, in Stock and Watson 1989's coincident index model famously missing the 1990-91 recession when the authors used it in real-time(Hamilton 2011, Katayama 2013). This downside arises in part because not all turning points are presented alike in the data, posing a vulnerability when future recessions do not mirror historic

[2]Usually defined as the 10-year minus 3-month US treasuries

[3]Because models in the literature generally produce probability-based outputs, it is up to the analyst to decide on a threshold value and/or decision rule to guide when to officially recognize a given period as recessionary

behavior across a limited set of predictors[4]. Finally, numerous past forecasting studies have conflated in-sample fit with out-of-sample performance (an error made mathematically explicit in Hansen 2008) while not adequately evaluating the latter. This can be especially problematic from an academic perspective, because look-ahead bias - i.e., basing predictions on data not available to the analyst at the time of model estimation - is easy to encounter under *ex post* out-of-sample testing, and can produce spurious results that would not hold under more rigorous evaluation designs.

One response to the dilemmas above has been increased focus on evaluating and furthering the development of nowcasting models. While recent innovations in nowcasting have gained media attention for providing advanced estimates of GDP and inflation[5], their practical use dates back to at least the 1980s (Fitzgerald and Miller 1989). In the business cycle literature, such methodologies have conventionally taken one of two forms: logit/probit regressions[6], and Markov-switching models. Efforts to improve real-time recession predictions have been generally focused on analyzing which variables hold the most predictive value under a given model, and making incremental improvements to existing methodologies.

Recently, predictive models developed in the artificial intelligence and statistical learning disciplines have shown promising results in a variety of applications, e.g., online shopping recommendations, securities trading, self-driving vehicles, etc. As a result, researchers from a number of different quantitative fields have adapted machine learning techniques to address data-related problems in their respective domains. While some progress has been made by economists in this regard, considerable work remains in understanding the extent to which

[4]One present concern, for example, is that following the Federal Reserve's unprecedented era of zero interest rate policy (ZIRP) in response to the 2007-09 financial crisis, the yield curve cannot technically invert without short duration treasury rates being raised materially above the zero lower bound

[5]See, for example, the Atlanta Federal Reserve's GDPNow model, and MIT's Billion Price Project

[6]Logit and probit models are often grouped interchangeably because their underlying cumulative density functions (logistic for the former, and normal for the latter) tend to produce similar probability outputs

the current set of standard econometric techniques can be improved upon or expanded. Consequently, this paper seeks to address the following:

1. How does the performance of conventional econometric models compare with those seen in the machine learning literature for nowcasting US recessions?

2. What advantages, if any, might machine learning techniques offer with respect to modeling nonlinear behavior across the business cycle?

3. Can ensembles of models produce out-of-sample performance surpassing that of each model individually?

The remainder of this paper proceeds as follows. Section II provides an overview of developments in the business cycle forecasting and nowcasting literature before considering recent work in applying machine learning techniques in this setting. In section III, treatment is given on the statistical and intuitive properties of the machine learning models evaluated in this study. Namely, this includes k-nearest neighbors (kNN), support vector machines (SVM), naive Bayes, random forests, gradient boosted trees, and artificial neural networks (ANN). Section IV features model performance comparisons for an out-of-sample test period over 1/1980-8/2014 and discusses results.

II Literature Review

Over the past few decades, a majority of studies in the business cycle prediction literature have implemented one of two models: logit/probit regressions, and/or Markov switching models. While the origins of logit/probit models can be traced back to the late 1800s (Cramer 2002), Markov-switching methodologies were introduced to the field in J.D. Hamilton's 1989 article, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." In it, Hamilton detailed a nonlinear technique for modeling economic regime changes as an unobserved first-order Markov process, with transition probabilities from one regime to another inferred via observed macroeconomic data[7]. To illustrate the model's potential, Hamilton estimated US expansion and recession dates based on inferences from current and lagged GNP data to arrive at a dating chronology independent of NBER. Using data from 3/1951-12/1984, Hamilton's model produced cyclical peak and trough dating estimates to within $\pm$ 2 quarters of the official NBER dates. As such, Hamilton 1989 provided an initial demonstration for the inherent potential of Markov-switching models to classify business cycle turning points in real-time.

Despite promising performance, Hamilton's originally-specified model left open a number of issues which have since been addressed through additional developments. First, because GNP and GDP are published on a quarterly basis and undergo considerable revisions following initial releases, a natural response was to estimate the model using monthly economic variables. Filardo 1994 took this approach, estimating a Markov-switching model on monthly industrial production data while setting transition probabilities from one economic regime to another as fixed across time (as was the case in Hamilton 1989). Despite a larger set of observations (i.e., 1/1948-8/1992), Filardo found that recession probabilities from the originally-specified model failed to cross the 50% threshold for five of the nine recessions tested in-sample. This disappointing performance, however, was ameliorated by instead using Stock and Watson 1989's composite leading indicator index as the explanatory variable, and allowing regime transition probabilities to change across time. Filardo concluded the latter model to have produced both the most optimal in-sample fit (measured according to Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC)), while also yielding competitive results out-of-sample from 1/1989-12/1991.

Additional research has sought to reconcile the Markov-switching model with

[7]While Hamilton was the first to apply this technique in a macroeconomic time series setting, Markovian approaches for modeling unobserved regime changes based on observed data were originally developed in the 1960s (Kouemo 2011) and applied to speech recognition in the 1970s (Rabiner 2015). In the statistical learning literature, these models are known as Hidden Markov Models (HMM)

Burns and Mitchell's observation that recessions are widely dispersed across the economy and can be inferred from the co-movements of leading and coincident indicators during cyclical turning points (see Chauvet 1996; Diebold and Rudebusch 1996; Layton 1996; Chauvet and Hamilton 2005; Chauvet and Piger 2008; Chauvet and Senyuz 2012). In recognition of this trait, independent developments from Diebold and Rudebusch 1996 and Chauvet 1998 extended the Markov-switching model through the inclusion of dynamic factors, a technique for identifying common behavior across a large set of economic indicators to generalize their co-movements into a small set of variables known as factors (Stock and Watson 2010b). As a result, additional indicators available at monthly and daily frequencies could be used to estimate models with further lead time ahead of NBER announcements, while at the same time maintaining model parsimony (Diebold and Rudebusch 1996; Chauvet 1996; Aruoba, Diebold and Chiara 2008). Today, incorporating a dynamic factor or first principal component variable with autoregressive lags to estimate business cycle turning points is considered standard practice in the Markov-switching literature.

Aside from efforts to extend the Markov-switching model's practical use, econometricians have also evaluated its performance in real-time regime classification. Chauvet and Piger 2008 took this approach by comparing a dynamic factor Markov-switching model (DFMS) based on NBER's four monthly coincident indicators (nonfarm employment, industrial production, real manufacturing and trade sales, and real personal income excluding transfer payments, henceforth referred to as the "Big 4 NBER indicators") to the Harding and Pagan algorithm (MHP) - a nonparametric technique for detecting cyclical turning points according to local minima and maxima present in each indicator. Using a data set with values as they were initially published prior to revision, the researchers observed superior results from the DFMS model, with expansion and recession dates falling "within one month, and never more than two months, from the corresponding NBER date" (Chauvet and Piger 2008).

In arriving at these results, Chauvet and Piger incorporated a "two-step" approach for detecting cyclical peaks. Specifically, this entailed waiting for the probability of recession to breach an 80% threshold and remain above this line for three consecutive months for a new recession to be recognized by the model. Thereafter, the initial month this threshold was breached would be defined as a peak date in the business cycle. Given these measures, the DFMS model was found to produce no false signals, though with the caveat that cyclical peak dates tended to be identified after an NBER announcement was made. On the other hand, troughs were called on average 8 months ahead of NBER, signaling a potential application for economic agents more interested in the end of a recession than the beginning (e.g., value-oriented investment managers).

Parallel to the development and evaluation of Markov switching models, econometricians have also published a number of studies evaluating logit/probit models in predicting recessions. Estrella and Mishkin 1996, for instance, found that a probit model based on the yield curve outperformed models with more expansive factor-based indices of leading and coincident indicators in forecasting out-of-sample recessions 2-6 quarters ahead, while producing competitive (albeit inferior) results at the 1 quarter horizon.

Dueker 1997 compared static, dynamic[8] and Markov coefficient-switching probit models based on the yield curve to forecast recessions 3-12 months ahead, and concluded that the lattermost tended to produce only marginal improvements in log-likelihood for all forecast periods except the 12 month horizon. However, because the author appeared neither to control for look-ahead bias nor implement out-of-sample results, these results should be taken with a generous level of skepticism.

Birchenhall, Jessen, Osborn and Simpson 1999 compared logit and Markov switching models for 1 and 3 month forecasts, and found "substantially more accurate business-cycle regime predictions" for the former. However, their out-of-sample test period included only one recession (1990-91), and as such, did not provide an adequate period for performance comparison across models.

[8]Dynamic logit models feature autoregressive recession lags as predictors and allow the lag order to evolve through the iterated forecasting period

Kauppi and Saikkonen 2008 analyzed static and dynamic probit models based on the yield curve to forecast both in and out-of-sample recessions 1-6 quarters ahead, with the former producing superior results on 1-4 quarter horizons while the latter generated "at least" equal performance for longer horizons.

Rudebusch and Williams 2008 compared yield curve-babsed probit models to recession probability surveys given by the Society of Professional Forecasters (SPF), and found the former to outperform the latter on all horizons beyond one quarter. The authors considered this result especially "puzzling" given that economists' interest in the yield curve as predictive of future recessions emerged around the 1980s.

Katayama 2010 examined the use of univariate, bivariate and trivariate logistic regression models to predict out-of-sample recessions 6 months ahead, while also considering the effect of different cumulative density functions (CDFs) on prediction accuracy. Similar to results from previous predictor comparison studies, the author found that a trivariate model based on the yield curve (given here as the 10 year Treasury minus Federal Funds rate), 3 month change in S&P 500, and nonfarm employment offered the best overall performance with respect to the logged probability score (LPS) and quadratic probability scores (QPS)[9]. Notably, Katayama also found superior predictive performance through the use of Laplace and Gumbel CDFs (as opposed to more commonly-used normal and logistic CDFs used in conventional probit and logit models, respectively).

Hamilton 2011 surveyed historical efforts at forecasting US recessions, as well as more recent developments in nowcasting methodologies. Similar to Hamilton 1989, a Markov-switching model with GDP as the explanatory variable was implemented for classifying business cycle regimes. Under this approach, recession starting dates were recognized on average one month within their announcement

[9]For observations $t = 1, \dots, T$, $S_t = 0$ if NBER expansion, $S_t = 1$ if NBER recession, and $\hat{p}_t :=$ model recession probability at time t, LPS is defined as $\frac{1}{T}\sum_{t=1}^{T}(S_t - \hat{p}_t)^2$, and compares both model accuracy as well as probability calibration. Similarly, QPS is defined as $-\frac{1}{T}\sum_{t=1}^{T}[S_t \log(\hat{p}_t) + (1 - S_t)\log \hat{p}_t)]$, and is analogous to the mean-squared error for binary classification accuracy.

by NBER, while expansion starts were identified on average three months prior to announcement. Hamilton also considered Markov models based on multiple monthly indicators (e.g., the a dynamic factor of the Big 4 NBER variables as in Chauvet and Piger 2008), but reasoned that while higher frequency data could provide "additional useful information" in classifying regimes, caution should be warranted under real-time applications given monthly indicators' more volatile probability outputs relative to models based on quarterly data.

Owyang, Piger and Wall 2013 found that a probit model combining both national and state-level predictors through Bayesian Model Averaging (BMA) produced "substantially" better out-of-sample results than those based solely on national data for 0-1 month recession predictions, though results across models tended to converge at the 2-3 month horizon. Given that some states' business cycles tend to lead national data while others lag, and since this relationship might reasonably be expected to evolve across time, the study represented a novel approach at incorporating a large set of correlated variables with little *a priori* knowledge on the predictive value of each variable individually[10].

Fossati 2016 compared the performance of static and dynamic probits to Markov switching models in nowcasting business cycle regimes, with separate models for "small data" and "big data" factors[11]. Implementing an out-of-sample test based on real-time data, the economist found that static logit and Markov switching models based on the "small data" factor produced the most attractive LPS and QPS. However, Fossati concluded that the "best" model should be decided by the analyst's individual cost function for accuracy versus timeliness. In this regard, static probit models performed optimally with respect to detecting cyclical peaks and troughs sooner, but at the expense of more volatile probability outputs (and in turn, false positives). Markov switch-

[10]Other methodologies for handling large sets of multicollinear predictors have also been implemented by econometricians. Ridge regression, one such example, imposes a shrinkage parameter on closely-related variables to avoid model overfitting. See, for example, Exterkate, Groenen, Heik and Dijk 2013

[11]Fossati defined the "small data" variable as a dynamic factor based on monthly growth in the Big 4 NBER variables, and the "big data" factor as the first principal component of "a balanced panel of 102 monthly US macroeconomic time series" (Fossati 2016)

ing models, for their part, tended to lag in calling recessions, but produced probabilities more heavily clustered around 0 and 1. Finally, in contrast to Kauppi and Saikkonen 2008, Fossati found dynamic probit models to yield inferior performance both in and out-of-sample.

With respect to applying techniques from machine learning to classify business cycle regimes, the econometric literature has been more sparse. While researchers have considered learning techniques in relative isolation, there has not been, to the author's knowledge, any comprehensive analysis comparing multiple different methodologies. With that said, articles analyzing machine learning methodologies in economics date back to the 1990s, with publications increasing in the last decade.

Vishwakarma 1994 analyzed a four-layer neural network using the Big 4 NBER variables to date US business cycle regimes from 1965-1989. While this model produced a business cycle dating chronology closely matching that of NBER, the study published in-sample results only. (While Vishwakarma mentioned conducting a short out-of-sample test from 1/1990-12/1992, these results were not published in the study.) Consequently, the paper more so illustrated neural networks' potential use in retroactively dating business cycles rather than classifying current regimes or forecasting turning points.

Gonzalez 2000 also evaluated the use of neural networks in macroeconomic time series, but with a focus on regression[12]. To illustrate their potential, Gonzalez conducted an out-of-sample test to predict next-quarter growth in Canadian real GDP from 3/1996-6/1998, and compared the predictions to those from a linear regression model developed by the Canadian Department of Finance. Both models used a basket of leading and coincident indicators, including employment growth, consumer confidence, long-duration interest rates, and the government's budgetary balance as a percentage of GDP. To assure model parsimony and deter overfitting, Gonzalez implemented a single hidden layer with two units in the neural network topology. (This setup was chosen according to results from

[12]Like many machine learning models, neural networks can be used for both recession and classification purposes

cross-validation.) In the forecast period, the neural network model produced a mean-squared error (MSE) of 0.0774, compared with the regression model's 0.1295. In discussing the neural network's 40% reduction in MSE, Gonzalez concluded that significant nonlinear relationships between the set of predictors and real GDP were likely present. However, due to their slower computation times, trial-and-error-based protocol for optimizing network topology, and "black-box" nature of producing forecasts without straightforward interpretations, Gonzalez cautioned that neural networks should be viewed a complement (rather than replacement) of existing linear forecasting models.

Qi 2001 evaluated neural networks in forecasting US recessions 1-8 quarters ahead, and was (according to the author) only the second study to apply the model to the business cycle literature after Vishwakarma 1994. The author conducted forecasts using a similar observation period, forecast horizon and variable selection as Estrella and Mishkin 1998 to ensure results were comparable[13]. The study's universe of predictors under consideration included "interest rates and spreads, stock price indexes, monetary aggregates, individual macro indicators, [and] indexes of leading indicators, both by themselves and in some plausible combinations" (Qi 2001). Out-of-sample testing was conducted from 3/1972-3/1995[14]. For predicting turning points one quarter ahead, a univariate model based on Stock and Watson 1989 leading indicator index was ranked highest, followed closely by stock returns for the NYSE and S&P 500. Similar to other studies, Qi found the yield curve to be the single best indicator of future recessions at the 2-6 quarter horizon, but found that further performance improvements were possible by featuring additional variables. Tellingly, the set of "best" predictors was found to evolve through the decades, illustrating the potential for methodologies that can extract a predictive signal from a large set of macro indicators.

Inoue and Killian 2004 considered economic forecasting models where the set

[13]Despite methodological similarities, Qi 2001 evaluated predictive accuracy based on mean-squared forecasting error, while Estralla and Mishkin 1998 used pseudo-R-squared

[14]The first observation in the data set was 6/1967

of potential predictors is considered "large", albeit remains smaller than total observations. Under situations where the universe of eligible variables is large and contains a number of highly-correlated indicators, discerning which ones hold the best predictive value out-of-sample can be computationally-intensive. On the other hand, depending on the model used, including the entire set of predictors might lead to overfitting. To address this dilemma, Inoue and Killian applied a machine learning ensemble technique known as bootstrap aggregation ("bagging") to function as a variable selection method for such scenarios[15]. In addition to reducing the possibility that a given model overfits the data, bagging can also assist in reducing the variance of out-of-sample predictions; a more rigorous treatment of this attribute can be found Breiman 1996.

In addition to modifying Breiman's original bagging technique for selecting predictors in a regression setting, Inoue and Killian also extended the method to time series applications by incorporating a "block-of-blocks" bootstrap. In time series data, temporal dependencies (i.e., autocorrelation) often exist in the data. The block-of-blocks bootstrap attempts to reconcile this issue by dividing the original sample into an initial set of non-overlapping blocks of consecutive observations of size n, and then subdividing each of these block into sub-blocks of size $k < n$. Resamples are generated by drawing with replacement the first-level blocks of size n from the original set of observations, and then further sampling with replacement a sub-block from each block drawn. Blocks of observations are then concatenated together in the order they were drawn to form a new set of observations.

To test the technique's empirical validity in economic time series analysis, Killian and Inoue evaluated a series of regression models to forecast CPI inflation 1 and 12 months ahead based on an out-of-sample test period from 8/1983-7/2003[16]. For their benchmark model, the researchers implemented an dynamic autoregressive model with up to 12 lags of inflation used to predict fu-

[15] As it was originally proposed in Breiman 1996, bagging involves estimating a given model on the original data as well as a large set of bootstrap resamples, and then averaging the results to form the final model

[16] Observations started from 4/1971

ture values[17]. For their bagged regression model, Killian and Inoue considered 26 macroeconomic variables in addition to the lagged inflation terms. To arrive at this model, they proceeded according to the following steps:

1. From the unrestricted regression model, analyze each variable's statistical significance as being non-zero according to a two-sided t-test

2. Remove predictors that fail to show significance at the 5% level to form the variable-restricted model

3. Perform 100 resamples of the original observations based on the block-of-blocks bootstrap

4. Estimate the parameters of the variable-restricted model on each of these 100 resamples

5. Take an average of these 100 models to form the final bagged forecasting model

To ensure a thorough evaluation at each step in the process, Killian and Inoue measured the accuracy of the benchmark lagged CPI model compared with the unrestricted model from step 1, the variable-restricted model from step 2, and bagged model from step 5. Except the benchmark, each model was estimated with a fixed number of lagged terms for the economic indicators, with the range set at 1-6. At both 1 and 12 month horizons, the bagged regression model produced the lowest root mean-squared error (RMSE) in a majority of the lag orders tested. In some cases, the performance improvement was dramatic - e.g., independent of the lag order specified, the highest ranked bagged model for forecasting inflation 12 months ahead produced an approximate 20% reduction in RMSE over the highest ranked unrestricted model.

Ng 2013 evaluated boosting - a machine learning ensemble technique similar to bagging - in order to improve the out-of-sample performance of logit models

[17] AIC was used as the decision criteria for determining the lag order at each iteration in the forecast

in predicting recessions. In its most basic form, boosting is used to improve the performance of a group of "weak" models (defined as outputting predictions only slightly more accurate than chance) through an iterative process as follows:

1. Take a sample of the data, with each given observation assigned a uniform probability ("weighting") of being drawn
2. Find a "weak" model from the set of candidates that minimizes the training classification error ϵ
3. Reassign weightings such that observations that were misclassified by the previous model are given a higher probability of being sampled in future rounds
4. Repeat steps 2-3 until ϵ converges
5. Assign a final model by taking a weighted-vote of each "weak" model's classification output based on accuracy

After modifying this approach to address time series-specific considerations (e.g., autocorrelation), Ng analyzed univariate logit models based on a universe of 132 monthly economic indicators to assess their their predictive utility for 3, 6 and 12 month ahead recession forecasts from 9/1986-12/2011[18]. In line with previous studies on the matter, Ng found that any one variable's predictive value tended to evolve throughout time, lending to the observation that each business cycle presents unique behavior in the data. With that said, indicators generally considered predictive in previous analyses, e.g., the yield curve, were found to hold value in Ng 2013, although mining-based employment growth stood out as an unusual candidate.

Berge 2014 compared four different model combination techniques for predicting recessions at the 0, 6, 12, 18 and 24 month horizon. This included an equally-weighted model average, a BMA model, and linear and nonlinear boosting models. Forecasts were based on an out-of-sample period from 1985-2013,

[18] The starting observation in the data was 3/1963

with initial observations starting in 1970. Similar to other studies summarized above, Berge found that real economic variables, e.g., the NBER Big 4, held a majority of their predictive value under short-term horizons, e.g., the 0 and 3 month horizon. On the other hand, financial variables, e.g., the slope of yield curve and corporate credit spreads, standing ahead of the group for predicting recessions 6-24 months out. For the combination techniques, a nonlinear boosted model was the most accurate model at every horizon tested.

Giusto and Piger 2016 compared a machine learning technique known as learning vector quantization (LVQ) to the dynamic factor Markov switching model for nowcasting recessions in real-time. Similar to Chauvet and Piger 2008, the analysis was based on a real-time data set for the Big 4 NBER variables as they initially released. Out-of-sample testing was conducted from 12/1976-7/2013, with 2/1967 as the first observation in the data. The LVQ model predicted business cycle peak and trough dates that were, on average, within 0.8 months of official NBER designations. Moreover, peaks and troughs were detected 134 and 234 days, respectively, after their initial occurrence. The DFMS model, on the other hand, identified cyclical peaks on average one month after their official announcement by NBER, while detecting troughs on average 249 days within their initial occurrence. (For their part, NBER's Business Cycle Dating Committee announced recession and expansion start dates on average 224 and 446 days following their actual realization.) Consequently, the study represents the potential of LVQ in particular, and machine learning approaches in general, for real-time recession nowcasting.

III Model Descriptions

i Choice of Models under Consideration

This paper evaluates the out-of-sample performance of conventional econometric models to more novel machine learning approaches for nowcasting recessions from 1/1980-8/2014. The former group includes logistic regressions and

Markov-switching models, while the latter considers k-nearest neighbors, support vector machines, naive Bayes, random forests, gradient-boosted trees and artificial neural networks.

The selection criteria for machine learning techniques was based on results from Caruana and Niculescu-Mizil 2006, which evaluated ten families of machine learning models on 11 unique cross-sectional data sets to gauge out-of-sample performance. Taking the mean from 8 different metrics to gauge predictive accuracy, Caruana and Niculescu-Mizil observed the most compelling results from bagged trees, random forests and neural networks.

ii Probability Calibration and Platt's Scaling

To further improve accuracy, Caruana and Niculescu-Mizil 2006 implemented a technique known as Platt's scaling, which seeks to calibrate a given model's classification probability according to the distribution of empirical results observed in the data set. Stated otherwise, if a given economy is in a recession for 20% of the periods in the sample, a well-calibrated model should produce recession probabilities of 90-100% in 20% of its predictions[19].

Implementing Platt's scaling on time series data is straightforward, and proceeds by the following process:

1. Partition the set of total observations according to an in-sample estimation period and out-of-sample testing period.

2. Estimate the model of interest on the in-sample data in order to make predictions for the out-of-sample data.

3. Take the set of predictions and partition it into another set according to an in-sample estimation period p_i and out-of-sample testing period p_0.

4. Estimate a univariate logit model based on the original model's productions from period p_i in order to produce predictions new for p_0. The

[19] As probabilities are continuous over the unit interval, this requires "binning" the model's set of predictions, usually by increments of 0.1.

outputs from this logit model are defined as Platt-scaled.

Caruana and Niculescu-Mizil found that this technique was able to further improve out-of-sample performance beyond that observed in the initial models tested. However, results were not uniform. For models producing well-calibrated classification probabilities without additional measures, e.g., logistic regressions and neural networks, little-to-no additional improvement in accuracy generally took place[20]. However, models that tend to produce sub-optimal probability-calibrations, e.g., support vector machines, witness improvements in accuracy following Platt-scaling. (It is important to note that the eligible set of data for use in the Platt-scaling process is far greater in cross sectional models than it is for time series models, given the latter's data dependency structure.)

Consequently, in addition to gauging out-of-sample recession nowcasting performance from 1/1980-8/2014, this paper also analyzed whether Platt's scaling might further improve predictive accuracy for nowcasting recessions. As such, the in-sample estimation period for this exercise took place from 1/1980-12/2000, and the out-of-sample test period took place from 1/2001-8/2004. However, due to time constraints, results were not published.

iii Schools of Thought in the Machine Learning Literature

To the uninitiated, the set of models featured in the machine learning literature appears to be a disparate set of methodologies with no underlying organization structure. This, however, is not entirely the case. In his 2015 book, *The Master Algorithm*, computer science professor Pedro Domingos makes the provides the following different schools of thought in machine learning based on their underlying philosophical approaches:

1. **Analogizers**. Practitioners from this school of thought consider analogous reasoning to be the primary means by which individuals use past experiences to make assessments of the world. An example of this can

[20]The author readily acknowledges that using predictions from one logit model as the inputs for another is likely to be an exercise in redundancy.

be seen in financial markets, where investors might base interpretations of current events according to similar episodes from the historical record. The primary machine learning models developed from this approach - k-nearest neighbors and support vector machines - form the quantitative implementation behind analogous reasoning.

2. **Bayesians**. This approach incorporates Bayes' Rule, which bases a given event's probability on prior evidence and updates these probabilities as new information is encountered. Practitioners from this school have developed a number of machine learning models, including the naive Bayes model, Hidden Markov models (equivalent to the Markov-switching techniques discussed above), and Bayesian networks.

3. **Symbolists**. Here, the focus is on applying formal rules of logic to develop predictive models. As such, symbolists' primary model of choice is the decision tree, which makes produces binary classifications according to a series of nested if-statements. More advanced implementations of this idea include random forests and boosted trees.

4. **Connectionists**. Practitioners in this group develop predictive methodologies based on mathematical models of the brain. Consequently, their machine learning technique of choice is the artificial neural network. More recently, connectionists have gained prominence in the development of deep learning, an advanced class of neural network models featured in such applications as facial recognition and self-driving vehicles.

5. **Evolutionists**. Under this paradigm, evolutionary processes provide the framework for optimizing model performance. As a result, practitioners in this school focus on developing genetic algorithms to "evolve" competing sets of models into a final model. In contrast to the models discussed above, genetic algorithms are an optimization technique, not a predictive model.

Following the organization set out by Domingos, a discussion proceeds over the properties of machine learning models in this study based on their respective school of thought[21].

Analogy-Based Approaches

k-Nearest Neighbors

One of the most straightforward techniques in machine learning, k-nearest neighbors (kNN) are both intuitively simple and computationally fast. Consider a set of observations $t = 1, \ldots, T$ from a binary class y_t with predictive variables $x_{1,t}, \ldots, x_{n,t}$. For a new observation $x_{i,T+h}$, the model simply considers the classes of the nearest k neighbors in Euclidean space, with a majority vote determining how the new observation is classified[22]. To improve out-of-sample performance, a distance-weighted approach can also be taken, with historical observations that lie closer to the new data point $x_{i,T+h}$ given a higher weighting in the vote.

Despite its simplicity, one of the primary disadvantages of the kNN model is its susceptibility to deterioration in out-of-sample accuracy when the set of predictors x_i is large. This is because as the dimensionality of the variable space increases, the neighborhood of any given new data point becomes less dense with respect to observations in the sample. Consequently, implementing the kNN model typically requires pre-selecting predictors according to their performance in out-of-sample testing, or implementing a dimensionality-reduction technique such as principal components analysis.

Support Vector Machines

This SVM model also takes an intuitive approach to classifying new observations in a data set, though with an ingenious twist. As before, consider a set of predictors $x_{i,t}, \ldots, x_{n,t}$ and a binary class $y_t \in \{0, 1\}$. In the variable space, assume there exists a linear boundary separating observations in class 0 from

[21]Note that genetic algorithms were not featured in this study, and as such, further consideration is not taken

[22]Non-Euclidean spaces can also be implemented, though this is less common in the literature

observations in class 1. In this case, only those observations at the nearest margins of this boundary will determine where the boundary is placed in the space. These observations form sets known as support vectors. As such, the SVM model fits the linear separation boundary that maximizes the marginal distance between support vectors associated with observations from classes 0 and 1.

A natural response to the description above is the question of how SVMs address cases where the boundary between classes is not linearly separable. To this end, the model takes an ingenious approach. Using a technique known as the "kernel trick", the SVM maps the variable space into a higher dimension, and then fits the separating hyperplane that maximizes the margin between support vectors in this space. Thereafter, the space is mapped back to its original dimension, where the decision boundary now takes a non-linear form.

Given this method of fitting a maximum margin boundary, SVMs are generally more prone to overfitting than other techniques. Consequently, their use among machine learning practitioners has increased heavily since the model's initial popularized in the 1990s.

Bayesian Approaches

Naive Bayes

The naive Bayes model takes a first-principals approach to applying Bayes' Rule to binary classification. Given $x_{i,t}$ and y_t as defined above, Bayes' Rule provides the conditional probability that a new observation y_{T+h} is in class 1 as follows:

$$\Pr(y_{T+h}|x_i) = \frac{\Pr(y_{T+h})\Pr(x_i|y_{T+h})}{\Pr(x_i)}$$

$\Pr(y_{T+h})$ is known as the *prior* probability of the outcome. $\Pr(x_{i,t})$ is the probability of the predictor values. $\Pr(y_{T+h})\Pr(x_i|y_{T+h})$ is the conditional probability of x_i given y_{T+h}.

While implementing Bayes' for a multivariate set of predictors X would otherwise add considerable complexity to the simple formula given above, the naive Bayes model circumvents this issue by assuming that the probabilities for $\Pr(x_i), i = 1, \dots, n$ are independent of each other. Consequently, we have

$$\Pr(X|y_{T+h}) = \prod_{i=1}^{n} \Pr(x_i|y_{T+h})$$

thus simplifying the computation immensely.

While this is assumption of independence is a strong one to make (especially with respect to time series data), practical use of the model has shown it to provide competitive performance in domains it should otherwise fail, e.g., spam detection.

Symbolist Approaches

Decision Trees

Popularized by Leo Breiman in the 1980s, decision trees provide the foundation for two popular machine learning models - random forests and boosted trees. For continuous variables $x_{i,t}$ and binary y_t as defined above, classification-based decision trees operate by partitioning the space of predictors X according to a series of nested logical rules that take a tree-based form.

To illusrate this idea, let's consider a simple recession classification model that features only one predictor - monthly growth in nonfarm employment (NFP). Suppose we define the rule: if NFP $< 0\%$, classify the current period as a recession. Otherwise, classify it as an expansion.

Now consider a model that also incorporates monthly growth in industrial production (IP). To this, we add the following: if NFP $< 0\%$ and IP $< 0\%$, classify the period as a recession. Let's assume further that we think NFP is more important than IP for predicting recessions. We could further state that if NFP $< 0\%$ and IP $> 0\%$, the current period should be classified as a recession, while if NFP $> 0\%$ and IP $< 0\%$, the current period could be classified as an expansion. Finally, if NFP $> 0\%$ and IP $> 0\%$, classify the period as an expansion.

In practice, while the mechanics of decision trees are accurately summarized above, the actual approach for determining the value k such that $x_{i,t} > k$ forms a given classification rule differs. Starting from the node at the top of the decision tree and moving downward, the model forms each decision threshold by taking a "greedy" approach - i.e., for the above model where NFP is topmost, we would set k equal to the value that produces the most correctly-classified economic regimes. This process would then reiterate at each node down the decision tree. Consequently, the model partitions the variable space into a series of piecewise linear decision boundaries.

Despite their ease of interpretation, decision trees tend to suffer in practical use due to their proneness to overfitting the data. To address this dilemma, two prominent approaches have been taken in the literature - random forests, and boosted trees.

Random Forests and Boosted Trees

Random forests use bootstrap aggregation (an ensembling technique described above) along with a random assignment of variables for each node to train a group of decision trees (known as a forest) on a given set of observations before combining each model's prediction through a simple vote. This process operates as follows:

1. Generate a bootstrap sample of the data, and use this sample to estimate a decision tree
2. For each node in the decision tree, randomly select $j < n$ of the original predictors, and select predictor $\hat{j}$ that produces the lowest classification error in partitioning the variable space
3. Repeat steps 1-2 until some stopping criteria has been met

Boosted trees, for their part, operate on a similar basis as random forests, but implement a boosting algorithm in place of bootstrap aggregation. In addition, they do not feature the random variable selection process used in random forest models.

Connectionist Approaches

Artificial Neural Networks

One of the few machine learning models studied in the business cycle literature, artificial neural networks (ANN) are popular in a number of domains where model interpretation is given a low priority relative to predictive accuracy. ANNs feature a network of nodes known as neurons, with sets of nodes grouped by layer. In its most basic form, this model is equivalent to linear regression - i.e., explanatory variables are mapped to the input nodes, assigned weightings based on the minimization of some cost function, and then summed together to produce an output. More typically, however, ANNs feature at least one *hidden* layer. (See figure on the following page for illustration.)

Despite the nomenclature, the process of producing predictions from explanatory variables remains straightforward. After variables are mapped to the input layer, weightings to each variable are assigned as in linear regression. Weights of each variable are then mapped to the nodes in the hidden layer. From here, an activation function is used to map these weightings into a final output. This function usually takes the form of the logistic CDF as featured in logit models.

Due to the functional composition implied in the above process, parameter estimation for neural networks does not lend itself to straightforward calculus-based cost minimization techniques seen in linear regression. Consequently, a numerical optimization technique known as backpropagation is used. While the particular details of this approach are beyond the scope of this paper, a few points should be emphasized. First, the model is generally more computationally-intensive than other machine learning techniques. Second, one pitfall from this approach is that parameters may be estimated from local (rather than global) minima to the cost function. As a result, neural networks can be prone to unpredictable performance.

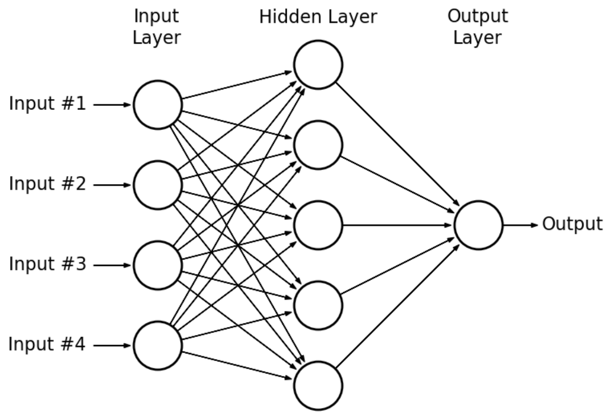

Figure 1: Diagram of an Artificial Neural Network.

IV Out-of-Sample Performance Comparisons

To gauge how the machine learning models detailed above compare to more conventional econometric techniques from the literature in nowcasting recessions, an out-of-sample test was taken from 1/1980-8/2014[23]. Variables were chosen according to their predictive performance in the literature (e.g., Berge 2013), and included monthly growth rates in nonfarm payrolls (NFP), industrial production (IP), real personal income excluding transfer payments (RPI), and private sector payrolls (PRIV). In addition, the annual growth rate in headline unemployment (UN), the yield curve (YLD), and the credit spread between Moody's BAA and AAA corporate bond yields (CORP) was also used in the universe of predictors. In line with previous studies, single period lags were taken for variables from the Big 4 NBER set, while 6, 9 and 12 month lags of the yield curve and corporate credit spreads were implemented.

In order to prevent look-ahead bias, assumptions were made on when to update the set of NBER recession dates. First, if an announcement was by NBER of the starting date of a given recession, the analyst would assume that all dates following the recession's start through the announcement date would

[23]The starting observation in the sample was 1/1960

also be considered recessions. This feature is based both off the persistence of economic regimes, as well as the lag given by NBER in announcing turning points. Consequently, the set of NBER dates classified by economic regime would be updated through the date of the most recent NBER announcement in the model. In instances where one year passed following the announcement of a business cycle trough, the analyst would assume observations from one year prior to the model's nowcast date were classified as expansions. This step was implemented to handle long periods of economic expansion in the data.

To allow for relative and absolute comparisons, two approaches were taken with respect to variable selection. First, each model was evaluated according to a standard set of variables - i.e., NFP, IP and their single period lags. While this approach was originally to be taken in recognition of past studies that have analyzed the Big 4 NBER variables, it was determined that real manufacturing and trade sales did not add any predictive value to the models evaluated (and in many cases, was associated with performance declines), while RPI only yielded marginal performance improvements in a subset of models.

After this initial evaluation, models were then tested based on the "best" variable subset for each model individually. For each model tested, the best variable subset is as follows:

- kNN: NFP, IP, IP.L1, PRIV, YLD.L6, YLD.L12
- SVM: NFP, NFP.L1, IP, IP.L1, UN, YLD.L9
- Naive Bayes (NB): NFP, IP, IP.L1, PRIV, YLD.L6, YLD.L12
- Random Forest (RF): NFP, NFP.L1, IP, IP.L1,
- Boosted Trees (BT): PRIV, PRIV.L1, IP, IP.L1, UN, YLD.L9
- Markov Switching Model (HMM): NFP, NFP.L1, IP, IP.L1
- Logit (LOG): NFP, NFP.L1, IP, IP.L1, RPI, YLD12

In line with previous literature, the first principal component was taken for the variables in the Markov switching model to maintain a univariate approach.

In addition to the above models, two basic combination approaches were taken. First, the classification output of each model was used to form a majority-vote model (MV). Second, the probability outputs of each model were averaged together to form an average-mean model (AP).

Results for the best variable selection tests are as follows:

	BT	HMM	KNN	LOG	NB	RF	SVM	MV	AP
Precision	0.84	0.91	0.98	0.83	0.84	0.93	0.89	0.95	0.87
Recall	0.93	0.77	0.95	0.96	0.96	0.95	0.89	0.98	0.96
F-1 Score	0.88	0.83	0.96	0.89	0.90	0.94	0.89	0.96	0.92
Area Under ROC Curve	0.95	0.88	0.97	0.97	0.97	0.97	0.94	0.99	0.97

Table 1: Best Variable Selection Test Results

Unsurprisingly, for the models based on a subset of the Big 4 NBER, results were less competitive:

	BT	HMM	KNN	LOG	NB	RF	SVM	MV	AP
Precision	0.71	0.91	0.84	0.82	0.74	0.79	0.90	0.86	0.82
Recall	0.84	0.77	0.88	0.89	0.91	0.89	0.84	0.89	0.89
F-1.Score	0.77	0.83	0.86	0.85	0.82	0.84	0.87	0.88	0.85
Area Under ROC Curve	0.89	0.88	0.92	0.93	0.93	0.93	0.91	0.94	0.93

Table 2: NFP, NFP.L1, IP, IP.L1 Variable Test Results

From the results, we conclude the following. First, despite their simplicity, logistic regressions stood up remarkably with respect to more novel techniques from the machine learning literature. Second, the simple k-nearest neighbor model outperformed every other model in the best variable selection test, while placing in the top 3 for models under the fixed variable test. Third, performance from the majority-vote model outperformed the top individual model in both tests, speaking to the potential use for model combination approaches based on a set of different methodologies. Finally, the performance from Markov switching models was relatively poor, while also producing the slowest computation times. Given that specifying the parameters for this model is a nontrivial exercise, caution is warranted on attempts to conclude the inferiority of the Markov switching model.

V References

Aruoba, S.B., Diebold, F.X., & Scotti, C. (2008). Real-Time Measurement of Business Conditions.

Berge, T., & Jorda, O. (2010). Evaluating the Classification of Economic Activity into Recessions and Expansions.

Berge, T. (2014). Predicting Recessions with Leading Indicators: Model Averaging and Selection over the Business Cycle.

Birchenhall, C.R., Jessen, H., Obsorn, D.E., & Simpson, P. (1999). Predicting US Business-Cycle Regimes.

Boldin, M.D. (1996). A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle.

Breiman, L. (1996). Bagging Predictors.

Burns, A.F., & Mitchell, W.C. (1946). *Measuring Business Cycles.*

Caruana, R., & Niculescu-Mizil, A. (2006). An Empirical Comparison of Supervised Learning Algorithms.

Chauvet, M. (1996). An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching.

Chauvet, M., & Piger, J. (2003). Identifying Business Cycle Turning Points in Real Time.

Chauvet, M., & Potter, S. (2005). Forecasting Recessions Using the Yield Curve.

Chauvet, M., & Hamilton, J.D. (2005). Dating Business Cycle Turning Points.

Chauvet, M., & Piger, J. (2008). A Comparison of Real-Time Performance of Business Cycle Dating Methods.

Chauvet, M., & Senyuz, Z. (2012). A Dynamic Factor Model of the Yield Curve as a Predictor of the Economy.

Cramer, J.S. (2002). The Origins of Logistic Regression.

Diebold, F.X., & Rudebusch, G.D. (1996). Measuring Business Cycles: A Modern Perspective.

Domingos, P. (2015). *The Master Algorithm.*

Dueker, M.J. (1997). Strengthening the Case for the Yield Curve as a Predictor of US Recessions.

Estrella, A., & Mishkin, F.S. (1996). The Yield Curve as a Predictor of US Recessions.

Exterkate, P., Groenen, P.J.F., Heij, C., & Dijk, D.V. (2003). Nonlinear Forecasting with Many Predictors Using Kernel Ridge Regression.

Filardo, A.J. (1994). Business-Cycle Phases and their Transitional Dynamics.

Fitzgerald, T.J., & Miller, P.J. (1989). A Simple Way to Estimate Current-Quarter GNP. Federal Reserve Bank of Minneapolis Quarterly Review, Fall, pg. 27-31.

Fossati, S. (2016). Dating US Business Cycles with Macro Factors.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The Elements of Statistical Learning.*

Giusto, A., & Piger, J. (2016.) Identifying Business Cycle Turning Points in Real Time with Vector Quantization.

Gonzalez, S. (2000). Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models.

Hamilton, J.D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle.

Hamilton, J.D. (2005). Regime-Switching Models.

Hamilton, J.D. (2011). Calling Recessions in Real Time.

Hansen, P.R. (2009). In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and its Implications for Model Selection.

Harding, D., & Pagan, A. (2003). A Comparison of Two Business Cycle Dating Methods.

Hechenbichler, K, & Schliep, K. (2004). Weighted k-Nearest Neighbor Techniques and Ordinal Classification.

Inoue, A., & Killian, L. (2004). Bagging Time Series Models.

Katayama, M. (2010). Improving Recession Probability Forecasts in the US Economy.

Kauppi, H., & Saikkonen, P. (2008). Predicting US Recessions with Dynamic Binary Response Models.

Kouemoe, G.L. (2011). History and Theoretical Basics of Hidden Markov Models. From Hidden Markov Models, Theory and Applications.

Kuhn, M. (2013). *Applied Predictive Modeling.*
Layton, A.P. (1996). Dating and Predicting Phase Changes in the US Business Cycle.

Malthus, T. (1820). Principles of Political Economy: Considered with a View to their Practical Applications.

Ng, S. (2013). Boosting Recessions.

Owyang, M., Piger, J, & Wall, H. (2013). Forecasting National Recessions Using State-Level Data.

Qi, M. (2001). Predicting US Recessions with Leading Indicators via Neural Network Models.

Rabiner, L.R. (2015). First-Hand: The Hidden Markov Model. Engineering and Technology History Wiki, 12 January 2015. Web. 31 May 2016.

Rudebusch, G.D., & Williams, J.C. (2008). Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve.

Simonde de Sismondi, J.C. (1819). *New Principles of Political Economy.*

Stock, J.H., & Watson, M.W. (1989). New Indexes of Coincident and Leading Economic Indicators.

Stock, J.H., & Watson, M.W. (2010a). Indicators for Dating Business Cycles: Cross-History Selection and Comparisons.

Stock, J.H., & Watson, M.W. (2010b). Dynamic Factor Models.

Varian, H. (2014). Big Data: New Tricks for Econometrics.

Vishwakarma, W. (1994). Recognizing Business Cycle Turning Points by Means of a Neural Network.

Zarnowitz, V. (1992). *Business Cycles: Theory, History, Indicators, and Forecasting.*