June 2010

Evaluating the Classification of Economic Activity into Recessions and Expansions *

Abstract

The Business Cycle Dating Committee (BCDC) of the National Bureau of Economic Research provides a historical chronology of business cycle turning points. This paper investigates three central aspects about this chronology: (1) How skillful is the BCDC in classifying economic activity into expansions and recessions? (2) Which indices of business conditions best capture the current but unobservable state of the business cycle? And (3) Which indicators predict future turning points best and at what horizons? We answer each of these questions in detail with methods novel to economics designed to assess classification ability. In the process we clarify several important features of business cycle phenomena.

- *JEL Codes:* E32, E37, C14
- *Keywords:* business cycle turning points, receiver operating characteristic (ROC) curve, Business Cycle Dating Committee of the National Bureau of Economic Research.

Travis J. Berge and Òscar Jordà
Department of Economics
U.C. Davis
One Shields Ave.
Davis, CA 95616

E-mail (Berge): tjberge@ucdavis.edu
E-mail (Jordà): ojorda@ucdavis.edu

*We thank Colin Cameron, Marcelle Chauvet, Puneet Chehal, Richard Dennis, Graham Elliott, John Fernald, Pierre Olivier Gourinchas, Robert Hall, James Hamilton, Fushing Hsieh, Kevin Lansing, Jeremy Piger, Glenn Rudebusch, Alan Taylor, Hal Varian, John Williams, the referees and editors of the American Economic Review and the American Economic Journal: Macroeconomics, as well as seminar participants at the Econometrics Workshop of the Federal Reserve Bank of St. Louis, the European University Institute, the Bank of Italy and U.C. Davis for useful comments and suggestions. Jordà is grateful for the hospitality of the Federal Reserve Bank of San Francisco during the preparation of this manuscript. Partial financial support was provided by the Spanish Ministerio de Ciencia e Innovación, grant SEJ-2007-63098.

1 Introduction

The Business Cycle Dating Committee (BCDC) of the National Bureau of Economic Research (NBER) was formed in 1978 to establish a historical chronology of business cycle turning points. The NBER itself was founded in 1920 and it published its first business cycle dates in 1929, although records are now available retrospectively starting with the trough of December 1854. Public disclosures of cyclical turning points are often made with more than a year's delay – the mission of the BCDC is not to serve as an early warning system to policy makers but to be a repository of the classification of economic activity for the historical record. Although other countries now have similar committees (such as the Euro Area Business Cycle Dating Committee of the Centre for Economic Policy Research founded in 2002), it is fair to say that the length of historical coverage and the experience of the BCDC have no equal.

If an economy's growth potential can be found by lubricating markets so that factors of production can be seamlessly combined to satisfy the desires of its people, it seems peculiar that an entire area of macroeconomics would be dedicated to business cycle research, let alone to keeping a simple binomial chronology of expansions and recessions. But the fact remains that cyclicality continues to be a salient feature of economic data, and even a casual observer will appreciate the relevance of understanding this cyclical behavior. For example, one would hope that economists could explain why unemployment could shoot up by four to five percentage points in less than a year, but then take the better part of a decade to shrink by as much. Understanding the short-term frictions that prevent efficient reallocation of resources in the face of shocks and force productive factors to sit idly for long periods of time explain the field's relevance.

Most models of aggregate economic behavior in the short-run are primarily based on log-linearized versions of dynamic stochastic general equilibrium (DSGE) models. These models generate stochastic processes in which the propagation of shocks is symmetric and moderately persistent. Newer models that incorporate search mechanisms into labor markets (à la Shimer, 2005, say) can generate more persistent behavior, but only in response to large initial exogenous shocks. But macroeconomic data suggest that there are periods in which a negative economic shock warrants no other action than waiting for the next positive shock to arrive, whereas other times that same shock is the precursor to a period of protracted economic stagnation. For this reason, a recent class of models incorporates the notion that economies behave differently depending on which state of the business cycle the economy is in when the shock occurs. Examples of such models include Farmer, Waggoner and Zha (2009) and Liu, Waggoner and Zha (2010).

A classification of economic activity into a binary indicator of recessions and expansions is therefore an effort to document statistical regularities that inform economists about features their models ought to explain. Moreover, decisions by economic agents appear to differ depending on the phase of the economic cycle. In a recession, policy-makers are more inclined to pursue aggressive fiscal policy (beyond that activated by automatic stabilizers), firms tend to postpone expanding their businesses, and workers tend to delay retirement. At a minimum, a business cycle chronology is an important research tool to study all of these issues.

The dating of business cycles does not amount to a mindless, mechanical, accounting exercise about when GDP growth (or employment, or any other indicator of economic activity) is observed to be negative. The BCDC's definition of a recession[1] states that:

> A recession is a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, real income, and other indicators.
>
> —*Determination of the December 2007 Peak in Economic Activity, December 11, 2008. Business Cycle Dating Committee of the National Bureau of Economic Research.*

[1] `www.nber.org/cycles/`

This definition, which harkens back to Burns and Mitchell (1946), implicitly recognizes that economic activity can be thought of as coming from a mixture of two distinct distributions (expansion/recession, say). Thus, even with an infinite sample, we cannot assign with certainty where a given observation belongs – just the probability that it belongs to either the expansion or recession distributions.

In view of the desire to document the different phases of the economic cycle, this paper asks three important questions: (1) How accurate is the taxonomy of expansions and recessions implied by the peak and trough dates recorded by the BCDC? (2) Because the BCDC releases are retrospective, which indicators best signal the current stage of the business cycle? And (3) which indicators predict future turning points best and at what horizons? These questions focus on evaluating classification ability rather than on providing new models of classification per se, research that we are currently conducting in a separate paper (Hsieh, Chen, Berge and Jordà, 2010).

Two features make evaluating a classification of cyclical economic activity particularly difficult. First, the true underlying state of the economy (expansion/recession) is never directly observable, even retrospectively. Second, *economic activity* itself is not characterized by a single, directly measured variable but by some combination of economic indicators that is not formally defined.[2] The first contribution of our paper is to provide a method to evaluate the first of these two features (the unobservability of the true state of the economy), while providing at least some progress in generating an index of economic activity.

Given this retrospective evaluation of cyclical classification ability, we then turn to the question of evaluating the classification ability of popular indices of business conditions, whose goal is to provide a more timely reading on cyclical economic activity. Here we find that the recently introduced Aruoba, Diebold and Scotti (ADS) index of business conditions[3] and the Chicago Fed National Activity Index[4]

[2] We thank Robert Hall for alerting us to the difficulties that the BCDC itself has in determining what is meant by economic activity.

[3] `www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/`

[4] `www.chicagofed.org/economic_research_and_data/cfnai.cfm`

(CFNAI), provide accurate signals about the current state of the business cycle. Lastly, we investigate the detection of future turning points using the components of the Conference Board's Index of Leading Indicators (ILI). Here we find that it makes little sense to construct a unique index to predict future cyclical behavior at all horizons: different components are optimal for classification at different horizons. Thus, a component that appears to be useless for some horizons (and which would receive a low loading in the ILI), can be quite useful at others (and thus deserve a higher loading). We provide out-of-sample evidence on direct predictive-classification ability up to 24 months into the future.

The methods that we use in this paper are mostly new to economics (there are some references in the credit risk literature, e.g. Khandani, Kim and Lo, 2010), although their earliest origin perhaps traces back to Peirce's (1884) "Numerical Measure of the Success of Predictions." Peirce's definition of the "science of the method" is the precursor to the Youden (1950) index for rating medical diagnostic tests, as well as the receiver operating characteristic (ROC) curve introduced by Peterson and Birdsall (1953) in the field of radar signal detection theory. The ROC curve methodology was quickly adopted into medicine by Lusted (1960) and is now a common standard of evaluation of medical and psychological tests (see Pepe, 2003 for an extensive monograph). The ROC curve approach has been adopted into fields as diverse as the atmospheric sciences (see Mason, 1982 for an early reference, as well as Stanski, Wilson and Burrows, 1989; and the World Meteorological Organization, 2000) and machine learning (see Spackman, 1989 for an early discussion). Recent applications to economics include, e.g. Jordà and Taylor (2010a, b).

Typical measures of forecasting accuracy for binary outcomes include the *mean absolute error* (MAE), the *root mean square error* (RMSE), and the *log probability score* (LPS), all of which rely on the specification of an underlying forecast loss function. However, a major advantage of the ROC curve is that it is not tied to a specific loss function as it itself is a map of the entire space of trade-offs (losses) for a given classification problem (we explain this issue in more detail below). Statistics based on the ROC curve therefore provide a non-parametric method for judging overall classification ability. Lastly, the new measures do not depend on the overall prevalence of recessions over the sample examined – this is important since recessions are observed only about 16 percent of the time. A rule that predicts every period to be an expansion will correctly predict expansions 84 percent of the time, a seemingly good number, but such a rule is clearly useless to policy-makers trying to head-off recessions since the rule has a 100% false positive rate (as it misses all the recessions). Our methods are set-up to explicitly recognize the policy trade-offs of these two rates.

2 Classification Ability: The ROC Curve

The methods that we use in this paper will likely be unfamiliar to most economists. The convention in economics is to investigate the marginal effect of a covariate on the probability that an outcome will be observed, and therefore consists of proposing a statistical model from which to generate predictions about the state of the economy, given a set of covariates. The covariates' predictive value can then be assessed with conventional inferential procedures. The loss functions associated with this predictive evaluation may vary, but if the specification of the model is a correct representation of the data generating process, one obtains unbiased estimates of the true model. However, when the statistical model is only an approximation, different loss functions result in different models and parameter estimates, and therefore possibly different conclusions about the usefulness of a particular economic indicator (see Hand and Vinciotti, 2003). The methods that we use here do not require that we construct specific models; we are able to separate the decision problem from the loss function. It is not that the loss function does not matter – it is crucial to determine what the optimal classification is for a given utility function over outcomes (see Elliott and Lieli, 2009). But when the utility trade-offs across outcomes are unknown, the methods we discuss provide more appropriate assessments of classification potential.

We now present our approach in detail by first discussing how to evaluate indicators taking the

BCDC's dating to be the true classification of business cycles. Later we will discuss the more nuanced question of how one can evaluate the BCDC's dating itself. Let $S_t \in \{0,1\}$ denote the true state of the economy, with 0 denoting that t is an expansion period and 1 a recession period instead. For the time being, assume that the BCDC can determine the value of this variable with 100% accuracy. Meanwhile, consider the index Y_t, which we require only to be ordinal. In our applications, Y_t will be a real-valued scalar. Y_t may denote a real-time probability prediction about S_t, a linear index, an index from a more complicated statistical model (e.g. a neural network estimator), or simply an observable variable (e.g. a leading indicator). The distinction is unnecessary for the methods we describe. Y_t together with the threshold c define a binary prediction *recession* whenever $Y_t \geq c$, and *expansion* whenever $Y_t < c$.

Associated with these variables, we can define the following conditional probabilities:

$$
\begin{aligned}
TP(c) &= P\left[Y_t \geq c | S_t = 1\right] \\
FP(c) &= P\left[Y_t \geq c | S_t = 0\right]
\end{aligned}
$$

$TP(c)$ is typically referred to as the *true positive rate*, *sensitivity*, or *recall rate;* and $FP(c)$ is known as the *false positive rate,* or *(1-specificity).*

The ROC curve plots the entire set of possible combinations of $TP(c)$ and $FP(c)$ for $c \in (-\infty, \infty)$. As $c \to \infty$, $TP(c) = FP(c) = 0$. Conversely, when $c \to -\infty$, $TP(c) = FP(c) = 1$, so that the ROC curve is an increasing function in [0,1] $\times$ [0,1] space. If Y_t is unrelated to the underlying state of the economy S_t and is an entirely uninformative classifier, $TP(c) = FP(c) \; \forall \; c$, and the ROC curve would be the 45^0 line, a natural benchmark with which to compare classifiers. On the other hand, if Y_t is a perfect classifier, then the ROC curve will hug the north-west border of the positive unit quadrant. Most applications generate ROC curves between these two extremes, although it is possible to imagine a "perverse" classifier that generates predictions that are worse than a coin toss (and therefore would generate a ROC curve that traverses the 45^0 diagonal). If such departures occur $\forall c$, it suffices to reverse the predictions from the classifier to generate a ROC curve strictly above the diagonal. Thus, since the abscissa is $FP(c)$ and c uniquely determines $TP(c)$, it is customary to represent the ROC curve with the Cartesian convention $\{ROC(r), r\}_{r=0}^{1}$ where $ROC(r) = TP(c)$ and $r = FP(c)$.

As an illustration, Figure 1 displays the ROC curve for an index of business conditions that we constructed to serve as a benchmark. The index is based on the number of printed news items with the word "recession" appearing in the LexisNexis database every month[5] since July 1970. The ROC curve displayed in the top panel of Figure 1 articulates the relative trade-offs in predicting recessions and expansions accurately. For example, correctly classifying 90% of all recessions results in a high rate of false positives (expansions incorrectly coded as recessions): 50%. By predicting recessions slightly less accurately – say, a true positive rate of 75% – the false positive rate would be cut in half to 25%. For completeness, the bottom panel of Figure 1 displays our index and the Google Trends[6] index for the word "recession" over the longest sample available from Google Trends.

In general, there may be different benefits and costs associated with making accurate predictions and errors; hence the overall utility of the classification can be expressed as (see Baker and Kramer, 2007):

$$\begin{aligned} U(r) &= U_{11}ROC(r)\pi + U_{01}(1 - ROC(r))\pi + \\ &\quad U_{10}r(1-\pi) + U_{00}(1-r)(1-\pi) \end{aligned} \tag{1}$$

where U_{ij} is the utility (or disutility) associated with the prediction i given that the true state is j,

[5] The index takes the raw counts of incidences per month, and adjusts for the trend in the number of news outlets included in the LexisNexis database over time and for seasonality. This index is similar in spirit to what Google Trends (visit www.google.com/trends) does to track the incidence of, e.g., influenza throughout the year. By tracking search activity on influenza related word searches, Google is able to provide a useful two-week ahead prediction of influenza incidence as reported by the Centers for Disease Control. We use our index in raw form—there is no model here—we just want to evaluate how useful is the index to classify the data into recessions and expansions based on the BCDC's chronology. We provide a more detailed description in the appendix.

[6] `www.google.com/trends`

$i, j \in \{0, 1\}$ and π is the unconditional probability of observing a recession in the sample. From the first order conditions in the maximization of expression (1), it is easy to see that

$$U_{11}\frac{dROC}{dr}\pi - U_{01}\frac{dROC}{dr}\pi + U_{10}(1-\pi) - U_{00}(1-\pi) = 0$$

or, rearranging

$$\frac{dROC}{dr} = \frac{U_{00} - U_{10}}{U_{11} - U_{01}}\frac{(1-\pi)}{\pi}. \tag{2}$$

That is, the optimum is that point where the slope of the ROC curve equals the expected marginal rate of substitution between the net utility of accurate expansion and recession prediction.

Underlying the classification problem is the view that the observations of Y_t reflect a mixture of two distributions. Specifically, let Z_t denote the observations of Y_t for which $S_t = 1$, with probability density function (*pdf*) given by f, and cumulative probability distribution (*cdf*) given by F. Similarly, let X_t denote the observations of Y_t for which $S_t = 0$ and with *pdf* given by g and *cdf* given by G. Then, the ROC curve can also be seen as a plot of $ROC(r) = 1 - G(F^{-1}(1-r))$ versus r, $r \in [0, 1]$, so that the slope of the ROC curve is

$$\frac{dROC}{dr} = \frac{g(F^{-1}(1-r))}{f(F^{-1}(1-r))}$$

that is, the slope of the ROC curve is the likelihood ratio between f and g. Hence, expression (2) relates the likelihood ratio between the expansion and recession distributions with the expected marginal relative utility from correct classification.

Given U_{ij}, $i, j \in \{0, 1\}$, one can therefore determine the *optimal operating point* as the threshold c^* that meets the equilibrium condition (2). Under the assumption $U_{ii} = 1$ and $U_{ij} = -1$ and $\pi = 0.5$, the optimal operating point maximizes the distance between $TP(c)$ and $FP(c)$, which is the well-known Kolmogorov-Smirnov statistic (Kolmogorov, 1933; Smirnov, 1939). Clearly the assumption $\pi = 0.5$ is violated for our analysis, and we do not know the values of U_{ij} that the BCDC uses. We revisit this issue in more detail in section 4.

A summary of all the trade-offs contained in the ROC curve and a commonly used measure of overall classification ability is the area under the ROC curve $(AUROC)$:

$$AUROC = \int_0^1 ROC(r)dr; \qquad AUROC \in [0.5, 1], \tag{3}$$

where it is clear that a perfect classifier has $AUROC = 1$ whereas a coin-toss classifier has $AUROC = 0.5$. A perverse classifier can generate an $AUROC < 0.5$ but then, by reversing the interpretation of the classifier's predictions from $S_t = 1$ when $Y_t > c$ to $S_t = 0$ (and vice versa when $Y_t < c$) the classifier would generate an $AUROC > 0.5$ so that for practical purposes an $AUROC = 0.5$ is the benchmark lower bound. This issue crops up in Section 5 and we show how it can be handled in practice there.

$ROC_A(r) > ROC_B(r)\ \forall\ r$ means that classifier A stochastically dominates classifier B, regardless of preferences. Although the $AUROC$ is the standard measure of classification for medical and psychological testing and for assessment of meteorological models, $AUROC_A > AUROC_B$ does not guarantee that $ROC_A(r) > ROC_B(r)\ \forall\ r$ (although $ROC_A(r) > ROC_B(r)\ \forall\ r$ does indeed imply that $AUROC_A > AUROC_B$). A direct test of the null hypothesis $ROC_A(r) = ROC_B(r)\ \forall\ r$ is available (see Jordà and Taylor, 2010b for a discussion of such a test and its distribution). However, rejection of this null does not always elucidate which classifier is preferred (if, for example, rejection came because $ROC_A(r) > ROC_B(r)$ over some range of r but $ROC_A(r) < ROC_B(r)$ for the complement range).

The $AUROC$ has several other convenient statistical interpretations. Green and Swets (1966) show that $AUROC = P[Z > X]$, where Z and X have been defined earlier. Therefore, a simple, nonparametric estimate of (3) is:

$$\widehat{AUROC} = \frac{1}{n_0 n_1} \sum_{j=1}^{n_0} \sum_{i=1}^{n_1} \left\{ I(Z_i > X_j) + \frac{1}{2} I\left(Z_i = X_j\right) \right\} \tag{4}$$

where $I(A)$ is the indicator function and is equal to 1 when A is true, 0 otherwise, and $n_k, k = 0, 1$ indicates the number of observations for the k^{th} state. The last term in (4) is a tie-breaking rule rarely needed when Y is a continuous index, as is the case in our applications. Bamber (1975) and Hanley and McNeil (1982) show that $\widehat{AUROC}$ is a two-sample, rank-sum statistic that can be reconfigured and reinterpreted as a Wilcoxon-Mann-Whitney U-statistic (Mann and Whitney, 1947 and Wilcoxon, 1945). Using empirical process theory, Hsieh and Turnbull (1996) show that under mild regularity conditions (described in detail in their paper and which we discuss in more detail below):

$$\sqrt{n_1}\left(\widehat{AUROC} - P[Z > X]\right) \xrightarrow{d} N(0, \sigma^2) \tag{5}$$

$$\sigma^2 = \frac{1}{n_0 n_1}\left[AUROC(1 - AUROC) + (n_1 - 1)(Q_1 - AUROC^2) + (n_0 - 1)(Q_2 - AUROC^2)\right]^{1/2}$$

$$Q_1 = \frac{AUROC}{2 - AUROC}; Q_2 = \frac{2AUROC^2}{1 + AUROC}.$$

For more details on the formulas for the variance see, e.g. Hanley and McNeil (1982), Obuchowski (1994), and Greiner, Pfeiffer and Smith (2000). The asymptotic normality result is very convenient because many hypothesis tests can be articulated using the familiar Wald principle (e.g. see Pepe, 2003). Bootstrap procedures are also available (see. e.g. Obuchowski and Lieber, 1998) although large sample approximations have been found to do well even in relatively small samples (again, see Pepe, 2003).

ROC curve methods provide formal assessment of classification ability: given the classifier Y_t, how well can it separate the classes associated with the true underlying states $S_t \in \{0, 1\}$. A non-parametric estimate of the $AUROC$ is easy to compute and its asymptotic distribution is Gaussian under general conditions so that inference against the null of no classification ability ($H_0 : AUROC = 0.5$) or comparisons of classification ability across classifiers, are straightforward (see Jordà and Taylor, 2010b for a detailed survey on other ROC-based testing procedures). The $AUROC$ is a two-sample, rank-sum statistic that compares the f and g densities implicit in the mixture distribution of Y generated by S for the basic problem of evaluating

$$P[Y_t \geq c | S_t = 1].$$

In the next section we consider a related evaluation problem: if Y_t is generated by an unobserved mixture process, we want to know whether the BCDC dates properly classify the data into each component of the mixture. What makes this evaluation problem difficult is that the true state of the business cycle is not directly observable.

2.1 Evaluating the BCDC's Dating

The BCDC dating has been taken by the profession and the public as an authoritative historical chronology of cyclical turning points. However, since the BCDC does not provide a mathematical or statistical algorithmic procedure that can be directly and formally evaluated, it is difficult to form a judgment about its quality.[7] Here we propose a possible solution to this problem. To be clear, we still have to deal with the issue of what is meant by *economic activity*. Our approach will be to investigate each of the economic indicators examined by the BCDC individually, as well as through a factor model. However, in the discussion that follows we often use GDP as our proxy for economic activity just to make the discussion flow more easily. Section 3 deals directly with economic activity broadly defined.

We begin by taking the view that economic activity can be approximately represented by a two-state

[7] Specifically, the latest public release of December 1, 2008 states that "Although the indicators described above are the most important measures considered by the NBER in developing its business cycle chronology, there is no fixed rule about which other measures may contribute information to the process in any particular episode."

mixture model so that an observation Y_t of, say GDP, could have come from a density f that characterizes recessions, or a density g that characterizes expansions. Moreover, it is natural to expect that the more extreme an observation (say an observation of 10% GDP growth) the more likely it is that it belongs to one or the other distribution (i.e. 10% GDP growth is more likely to belong to the expansion distribution g than say, 2% GDP growth is). For illustrative purposes, Figure 2 displays a kernel density estimate of the two distributions implied by the BCDC dating. The top panel displays the empirical mixture for the distribution of month-to-month (first log difference) annualized percent growth rate of real GDP whereas the bottom panel displays the mixture for the year-to-year (twelfth log difference) growth rate transformation instead. The sample of interpolated monthly data[8] begins March 1947/February 1948 and ends August 2009 (the difference in start dates reflects the monthly/yearly growth rate transformations). The mean of the recession distribution for the monthly/yearly transformation is -1.6/0.05% annual GDP growth whereas the expansion distribution is centered at 4.2/3.9%. The standard deviation of the recession/expansion densities for the monthly transform is 3.55/3.25 whereas it is 2.11/1.97 for the yearly transformation. Regardless of how GDP growth is calculated, it is apparent that there is a region of considerable overlap between the two distributions. Nevertheless, we will show that the BCDC has very high skill when classifying economic activity.

Therefore, think of the BCDC's dating as a filtered probability prediction $\widehat{S}_t$ of the unobservable, underlying class marker S_t. If $\widehat{S}_t$ were generated by a fair coin-toss, the resulting f and g densities of the mixture for Y would be identical to a null model in which Y is assumed to come from a non-mixture process. The $AUROC$ for this coin-toss classifier would be 0.5, the typical null. Instead, the more skill in the construction of $\widehat{S}_t$, the clearer the distinction between the implied mixture distributions; in fact, perfect classification will generate $AUROC = 1$. A considerable portion of the paper consists in evaluating potential classifiers of the true state of the economy based on the BCDC dates. Therefore,

[8] We use the linear interpolation method described in the BCDC's release of December 1, 2008 and which is described in more detail in the next section.

we use this metric to assess the skill of the BCDC prior to using the BCDC dates as "the truth." In particular, Section 4 is devoted to testing the BCDC against the coin-toss null as well as against alternative dating schemes based on two specifications of Hamilton's (1989) well known hidden Markov mixture model.

3 Preliminaries

There are four preliminary features of the analysis that affect any assessment of business cycle dating ability: (1) transitions into and out of a given state tend to be persistent; (2) different detrending methods produce phase shifts of the chronology that best sorts the data into expansions and recessions; (3) how does one define "economic activity;" and (4) different definitions of recession have varying classification skill.

3.1 Persistence

The persistence of each state can be illustrated with a novel concept that we introduce and that we call "the autoclassification function[9]" (ACF). The ACF is a plot of the $AUROC$ resulting from setting $Y_t = S_{t-h}$ for $h > 0$; that is, using past values of the state to classify the state in the current period. Note that if knowing the state in a previous period is not useful when classifying the current period, then the $AUROC$ will be 0.5 (not 0 as would be common in a typical autocorrelogram). Figure 3 displays the ACF of the recession indicator based on the BCDC's peak to trough dates (both peak and trough included) for horizons up to 12 months. The ACF clearly shows that knowledge of the state of the economy up to 8 months into the past carries an informative signal regarding the current state (at 9 to 12 months the $AUROC$ is virtually 0.5).

This persistence is undoubtedly useful to construct better classifiers: this fact is explored in much of

[9] We thank Colin Cameron for providing this suggestion.

the empirical business cycle literature e.g. in Neftçi (1982), Hamilton (1989) and Diebold and Rudebusch (1990). However, one may wonder whether this persistence substantially affects $AUROC$ inference. We believe that it may make traditional $AUROC$ variance estimates slightly more conservative (because the amount of persistence, for practical purposes, is very low). The reason is that the $AUROC$ compares the expansion and recession empirical densities directly: i.e., how often is $Z_i > X_j$, where Z_i denotes observations of Y_t during recessions, and X_j during expansions. The more "separated" these two distributions are, the closer the $AUROC$ is to 1, and the lower the variance of its estimate. Correlation reduces entropy and tends to increase estimation uncertainty for statistics based on a single density (such as a sample average). But correlation within state (expansion/recession) actually helps disentangle the two densities in the mixture more easily, thus producing a higher $AUROC$ and a lower variance estimate.[10]

3.2 Trends and Cycles

Next, we examine the effects of the detrending method on classification phase shifts, the second of the issues mentioned earlier. In a stable economy, growth rate transformations are perhaps least controversial and we examine two types:

1. Annualized percentage month-on-month growth rate transformation for any variable X_t is defined as:

$$x_t^m = 100 \times 12 \times \Delta \log X_t$$

2. Annualized percentage year-on-year growth rate transformation for any variable X_t is defined as:

$$x_t^y = 100 \times \Delta_{12} \log X_t$$

[10] We thank Fushing Hsieh for clarifying this point to us.

x_t^y is clearly a smoothed version of x_t^m but the latter most closely matches the timing of cyclical shifts.

In addition to these growth rate transformations, we briefly experimented with filters commonly used in the business cycle literature. However, this is done more for completeness than out of conviction because: (1) there is no consensus about the appropriate trend-cycle decomposition; (2) filtered trend estimates are sensitive to the sample used and may vary as the sample grows over time; (3) trends across indicators differ; and (4) common filtering methods introduce additional dynamic elements into the cyclical component.

Table 1 reports $AUROC$ estimates of cyclical GDP (for brevity we do not report estimates for other variables) using the two growth transformations x_t^m and x_t^y; a Hodrick and Prescott (1997) detrended series x_t^{HP}; a Baxter and King (1999) band-passed filtered series x_t^{BK} where the cycle is defined over frequencies between 6 and 32 quarters; and from a series detrended using the potential output series reported by the Congressional Budget Office x_t^{CBO}. The sample begins March 1947/February 1948 (for x_t^m/x_t^y) and ends in November 2007, the month prior to the last release available from the BCDC. Recessions are defined using the BCDC definitions as the period between a peak and a trough of economic activity, both included.

Virtually perfect classification ($AUROC = 0.98$) is achieved with x_t^y but this value is achieved by shifting the beginning and end of recessions by three months due to the implicit smoothing of such a transformation. x_t^m matches the BCDC's timing but its $AUROC = 0.89$, while high is statistically different than 0.98. Cyclical measures obtained with the HP filter, the BK filter and the CBO trend require an even bigger phase shift of 6 to 7 months while not attaining any significant $AUROC$ improvement over the simple growth transformations. For these reasons, we will focus on growth transformations for the remainder of this paper.

3.3 Economic Activity

The third facet in evaluating the BCDC's classification skill requires clarification of what is meant by "economic activity" (see the release statement quoted in the introduction). The BCDC cites five monthly indicators of economic activity: industrial production (IP); real personal income less transfers (PI); payroll employment (PE); household employment (HE); and real manufacturing and trade sales (MTS). It also considers two quarterly indicators: real gross domestic product (GDP); and real gross domestic income (GDI). Details on the sources and construction of each indicator follow the statements by the BCDC and are summarized in the appendix. We then take annualized month-on-month (x_t^m) and year-on-year (x_t^y) log differences to compute growth rates expressed in percentages, for clarity.

In addition, we constructed two aggregate indices of economic activity (one for x_t^m data and one for x_t^y data), based on factor analysis (using principal components as in Stock and Watson, 1989) and using these 7 indicators of economic activity. Figure 4 displays the main factor for the x_t^m data (top panel) and the x_t^y data (bottom panel) and we denote this factor "Principal Component Index" or PCI (or more specifically PCI^m*or* PCI^y depending on the transformation used). For each growth transformation, we note that the factor explains approximately 85% of the variability in the data and scree plots (not reported here) strongly indicate that a single factor is indeed sufficient. Interestingly, we also looked at the manner in which the factors load on each of the 7 indicators mentioned previously and found that each variable receives very similar weight so that movements in the factor are not dominated by any single series. Figure 4 shows that these factors also conform well with the business cycle and as we shall show momentarily, are almost perfectly classified by the BCDC.

3.4 Four Definitions of Recession

The fourth and last issue we discuss in this section refers to the definition of what is a recession. The BCDC produces a series of business cycle turning points for the U.S. economy that contains the month within which the day of a peak or a trough of economic activity occurs (see the BCDC's release of December 11, 2008). Each peak and trough month is therefore some mix of economic expansion and recession. It is generally accepted that trough months should be classified as recessions, but there is more ambiguity as to how peak months should be classified. The BCDC itself,[11] Chauvet and Hamilton (2005) and Wright (2006) define recessions as the period between a BCDC peak and a trough, including both the peak and trough months. We denote the series produced by this method *BCDC-PI* (for peak included). Rudebusch and Williams (2009) instead choose to date recessions by excluding peak months. We denote this rule *BCDC-PE* (for peak excluded). In addition, we consider two alternative and popular "rule-of-thumb" definitions of recessions. The first classifies recessions mechanically as any period in which GDP growth is negative. The other, quite popular with the media, calls a recession when there are at least two consecutive quarters of negative GDP growth. Following Rudebusch and Williams (2009) we will call the two series *R1* and *R2*, respectively.

Table 2 tabulates the salient features of each of these four recession definitions. The sample begins January 1947 and ends November 2007 since December 2007 is the latest BCDC date (a peak) available. During this period there are 114 months the BCDC classifies as recessionary (104 if peaks are excluded since there are 10 recessions during this period), a similar number to the 108 months found with the R1 rule but almost twice as large as the 63 months found with the R2 rule. The average BCDC-PI recession lasts just under a year on average (11.4 months), which represents about 16% of the sample. The R1 rule is very noisy relative to the other rules since recessions last less (4.5 months) but occur more frequently (24 times!). Conversely, the R2 rule is much more conservative (with half as many recessionary months detected) but misses the 2001 recession entirely.

Next, we use the ROC curve analysis to compare the classification skill of each of these four definitions relative to each of the coincident indicators mentioned in the BCDC's release of December 11, 2008 as

[11] `www.nber.org/cycle/`

well as our principal component index, PCI. Thus, Table 3 reports $AUROC$ estimates for each indicator using the two growth transformations (x_t^m and x_t^y) for dates lining up with the BCDC chronology (h = 0) and for the phase shift that would maximize the value of the $AUROC$ with a window h of ± 24 months, h^*. Generally speaking, either BCDC-PI or BCDC-PE are clearly superior to the R1 and R2 rules regardless of how the data is transformed. The month-on-month transformation matches the timing of the BCDC chronology more accurately in all the cases (with an occasional phase shift of one month here and there). In this case, BCDC-PE uniformly generates the highest $AUROC$ values of all four rules although the differences with respect to BCDC-PI are negligible. However, the year-on-year transformation usually attains significantly higher $AUROC$ values as long as one allows for phase shifts that range between three months for output related indicators, to six-months for employment related indicators. $AUROCs$ for the principal component indices are highest, coinciding with the BCDC dating for PCI^m ($AUROC$ = 0.94, 0.96) and with a 5/6 month shift PCI^y ($AUROC$ = 0.98). However, in this case there are no clear differences between BCDC-PI and BCDC-PE. For this reason and because BCDC-PI is the rule proposed by the NBER, from here on we use this timing convention when defining recessions.

4 The BCDC's Classification Skill versus Statistical Dating Rules

The BCDC's dating of business cycles is held as a natural benchmark against which competing methods of turning point prediction are evaluated. Even models in which the underlying state of the economy can be estimated independently of the BCDC's classification (such as the class of hidden Markov mixture models spawned by Hamilton's 1989 seminal work) evaluate their success when estimates of the smoothed state probabilities line up against the BCDC's peak-trough dates. This section turns this view point on its head and instead asks how well the BCDC dates compare to smoothed state probabilities available in Chauvet and Hamilton (2005) and Chauvet and Piger (2008).

We first consider a monthly interpolation of Chauvet and Hamilton's (2005) quarterly Markov-switching smoothed transition state probability index (henceforth the CH Index), which is readily available and transparent.[12] The two-state Markov chain specified in the model is based on GDP growth alone, and captures the underlying unobserved state of whether the economy is in expansion or recession. Chauvet and Piger (2008) also produce an index that captures aggregate activity, which we will denote as the CP Index.[13] CP first estimate a dynamic factor model in the vein of Stock and Watson (1989) using data on four monthly coincident variables: nonfarm payroll employment, industrial production, real manufacturing and trade sales, and real personal income less transfer payments. The common factor follows a two-state Markov process so the model results in smoothed recession probabilities.

In order to compare the performance of these two models to the BCDC dates, we first need to translate the state probabilities that result from these models into monthly zero-one indicators about the state of the economy. To create a binomial variable, we apply the simple rule-of-thumb that any period with a recession probability greater than a given threshold value c is classified as a recession. So that each model performs as well as possible, we search for the c that produces binomial dates that maximize the $AUROC$ relative to the Principal Component Indexes. For the CH Index, we allow c to vary over the space 0.3-0.9. However, to ensure that the CP Index produces a reasonable number of recessions, we placed an upper-bound of 0.6 on c when producing a binomial variable from the CP Index.[14]

The top panel of Table 4 displays the $AUROC$s associated with each of these two statistical-based recession indicators and the BCDC-PI dates and for each of the coincident indicators examined by the BCDC and our PCI indexes. The bottom panel repeats the exercise but allows for a phase shift h in the range of ± 24 months. We then report the maximum $AUROC$ achieved and the corresponding h^*.

[12] `www.econbrowser.com/archives/rec_ind/description.html`

[13] Available from Jeremy Piger's homepage at `www.uoregon.edu/~jpiger/`

[14] We remark that when we do not constrain c in this manner, the CP index would have produced three recessions between 1967 and 2007. In that case, the index achieves an AUROC = 0.997 based on the PCI – virtually perfect classification.

Broadly speaking, it is difficult to find any significant differences that are worth remarking on. Although CP is optimized over a broader collection of series, it does not seem to improve over CH (it seems to do best for year-on-year transformed data but worse for month-on-month transformed data). In any case, neither statistical procedure appears to improve very much over what the BCDC does. Looking at PCI, one would be unable to reject the null that all three methods produce equally competent classification.

5 Indices of Business Conditions

The analysis in Section 4 justifies that from here on we take the chronology of peaks and troughs of economic activity provided by the BCDC as a very reasonable barometer of the true state of the economy. However, because the BCDC releases are made with a lag of 12 to 18 months, in this section we investigate whether there are indicators of business conditions that may provide the public with a more timely signal. We follow-up on this question in the next section by investigating prediction of future turning points up to two years into the future.

For now, we investigate three popular indices of aggregate economic activity, plus the LexisNexis news-based indicator that we introduced in Section 2 as a benchmark. These are indices commonly used in the profession and are freely and publicly available. Two of the indices represent state-of-the-art approaches to measuring aggregate economic activity in real time. The Chicago Fed National Activity Index (CFNAI) is a monthly index constructed as a weighted average of 85 monthly indicators of national activity drawn from four broad categories: production and income; employment, unemployment and hours; personal consumption and housing; and sales, orders and inventories. The CFNAI corresponds to the index of economic activity introduced in Stock and Watson (1999). More details can be found in their paper and in the Federal Reserve Bank of Chicago's website.[15] The second index included is

[15] www.chicagofed.org/economic_research_and_data/cfnai.cfm

the Aruoba, Diebold and Scotti (ADS) Business Conditions Index maintained by the Federal Reserve Bank of Philadelphia. The ADS index is a new index designed to track real business conditions at very high frequencies. It is based on a smaller number of indicators than CFNAI, and the details about its construction can be found in Aruoba, Diebold and Scotti (2009) or at the Federal Reserve Bank of Philadelphia's website.[16]

The other two indices we investigate rely on information from market participants instead of attempting to measure economic activity directly. The first index is the manufacturing Purchasing Managers Index (PMI), which has been issued since 1948 by the Institute of Supply Management. The data for the index are collected through a survey of 400 purchasing managers in the manufacturing sector. The PMI is available at a monthly frequency (more details can be found at the Institute of Supply Management's website[17]). We also include the index that we introduced in Section 2 based on a standardized measure of the counts of printed news items containing the word "recession" in the LexisNexis academic database. This crude index is meant to provide a benchmark of comparison for the three other indices described above.

The results of this analysis[18] are reported in Figures 5 (for CFNAI), 6 (for ADS) and 7 (for PMI), each of which contains two panels (we remind the reader that the graphs for the LexisNexis index were already presented in Figure 1). The top panel displays the ROC curve (using the BCDC-PI recession dates discussed in Section 3) and the bottom panel the time series for the index. Both the CFNAI and ADS indices do very well with $AUROC$ values of 0.93 and 0.96 respectively, fairly close to near-perfect classification ability. The PMI index has an $AUROC = 0.9$, which is somewhat lower but PMI is a

[16] www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/

[17] www.ism.ws/ISMreport/content.cfm?ItemNumber=10752&navItemNumber=12961

[18] We evaluate these indices with the most recently available data vintage since real-time vintages are not available for a long enough period, nor are they available for all the variables that we consider. However, we do not think this is an important limitation - although data revisions can sometimes be considerable for a single variable (such as GDP), these changes affect the indices to a much smaller degree. Moreover, Chauvet and Piger (2008) show that data revisions do not seem to affect the actual dating of business cycle turning points. It would be interesting to look at real time data in more detail but this is left for further research.

narrow indicator for production rather than a broad-based measure such as CFNAI and ADS. As a benchmark, our LexisNexis index has an $AUROC = 0.81$, which is statistically inferior to any of the three indices considered. A more detailed investigation into the indices themselves revealed that out of the variables included to construct the ADS index, initial jobless claims alone has an $AUROC = 0.95$, which is considerably higher than any of the other variables and approximately the value attained by the ADS index itself. These results are summarized in the top panel of Table 5.

Before we conclude this section we make two observations. First, the classification ability of all the indices considered deteriorates very rapidly when used to predict turning points into the future: within a year, they are no better than a coin-toss at distinguishing recessions from expansions, as can be seen from the top panel in Table 5. Second, we calculated threshold values that would maximize the utility of the classification so as to check the values recommended by the different agencies that publish these data. For this purpose we make the working assumption that the benefits of hits equal the costs of misses in magnitude. Under these conditions, the optimal threshold can be determined from expression (1) as:

$$\max_{c} \left(2\widehat{\pi}\widehat{TP}(c) - \widehat{\pi}\right) - \left(2\left(1-\widehat{\pi}\right)\widehat{FP}(c) - (1-\widehat{\pi})\right)$$

where $\widehat{\pi} = 0.16$, the unconditional probability in the sample of an observation belonging to a recession. The resulting estimates of the optimal thresholds are for CFNAI, $c^* = -0.72$; ADS $c^* = -0.80$; and $PMI = 44.48$, which are somewhat lower than the values commonly used as rules of thumb, which for CFNAI and ADS is $c^\star = 0$ (although we point out that Chicago Fed posts a document suggesting that $c^\star = -0.7$ seems to provide a more accurate chronology of turning points) and for PMI is $c^\star = 50$. Of course, these estimates would vary under different assumptions about the relative utility of classification hits and misses. The bottom panel of Table 5 summarizes how observations since December 2007 up to October 2009 would be classified by these business conditions indicators and show that the end of the recession would be dated to be September 2009 for CFNAI (although the last release dipped slightly below the optimal threshold), July 2009 for ADS and June for PMI.

These dates conform well with recent statements made in the press. For example, *The Economist* on October 29, 2009 reports that "Robert Gordon, a member of this group [the Business Cycle Dating Committee], is confident that the recession, which began in December 2007, ended in June." Robert Hall, who chairs the BCDC, declared for *Bloomberg* December 4, 2009 that "The trough in output was probably some time in the summer." Alan Greenspan declared in *Meet the Press,* December 13, 2009 that the recession ended July 2009, possibly as early as June 2009.

6 Future Turning Points

The last of the three main questions we set out to investigate in this paper considers the ability to predict future business cycle turning points. In this section we focus on the components of the Conference Board's Index of Leading Indicators (ILI), a complete description of which is provided in the appendix. Throughout this section we maintain the working convention that the BCDC's chronology is the "gold standard" that these predictions should properly classify. Within this section, we accomplish two tasks. First we use ROC analysis to determine the relative classification ability of each individual component of the ILI over horizons ranging from 0 to 24 months in advance. Interestingly, we find considerable variation in classification ability across predictors and across forecast horizons. More specifically, we find that at some horizons, positive values of the predictor are associated with higher likelihood of recession, whereas at other horizons the association is with higher likelihood of expansion. This non-monotonicity is revealing because it suggests that parsimonious affine models will often lack sufficient texture to generate accurate predictions of the economic cycle, even a few periods into the future. Thus, the second task we carry out is a direct prediction-classification exercise and out-of-sample evaluation over several horizons.

6.1 The Conference Board Index of Leading Indicators: ROC Analysis

The Conference Board's Index of Leading Indicators includes ten individual components (see appendix for data sources and description). Several of these variables are meant to capture market or consumer expectations about future economic activity—for example, the S&P 500 stock market index and the Treasury debt yield spread between the 10-year T-bond and the federal funds rate (FFR) both speak about market expectations, while the University of Michigan consumer survey directly measures household expectations. The remaining variables—building permits for new housing units, average weekly hours in manufacturing, manufacturers' new orders, initial claims for unemployment insurance, and the index of supplies deliveries — are more direct measures or precursors of future economic activity.

Figure 8 displays the $AUROC$s across horizons $h = 0, 1, ..., 24$ for all ten leading indicators used by the Conference Board and using month-on-month (top panel) and year-on-year (bottom panel) growth transformations of all the variables except the 10-year-FFR spread. In the interest of readability, we break-up the indicators into two panels and suppress confidence intervals. Many indicators achieve $AUROC$ maxima at horizons very close to $h = 0$. Interestingly, however, these indicators then achieve minima at horizons between 12 and 18 months into the future. As we explained in Section 2, an $AUROC < 0.5$ means that we have a "perverse" classifier whose performance is worse than that of a coin-toss but whose reciprocal would have an $AUROC > 0.5$ and hence be useful in properly classifying the data. Consequently many of the indicators appear to have valuable information to forecast recessions at distant horizons as long as one flips the sign of the index.

For these reasons, Table 6 summarizes the maximum and the minimum $AUROC$ achieved by each indicator along with the month-on-month and year-on-year growth transformations, and reports the horizon at which these optima are achieved. In the interest of clarity, the minima are expressed in the usual $AUROC$ scale in the interval $[0.5, 1]$. Broadly speaking, the year-on-year transformation achieves considerably higher $AUROC$ values than the month-on-month transformation. Several indicators achieve their highest $AUROC$ values either contemporaneously or within the first couple of months with one notable exception: the 10-year T-bond-FFR spread's $AUROC$ is maximized 18 months in the future.

As an example of the behavior of a typical indicator, consider new orders for consumer goods. The maximum $AUROC$ is 0.93 (for the year-on-year transformation) and is achieved one-month ahead. However, the reciprocal of the index achieves an $AUROC$ of 0.71 when looking 22-months into the future, which is well above the 0.5 value of no classification ability. Moreover, recall that the ACF in Section 3 and reported in Figure 3 suggests that there is no classification information to be gained from past values of the state variable S_{t-h} for values of h beyond 8 months, making the value $AUROC = 0.71$ 22 months into the future all the more remarkable.

Many of the components of the ILI exhibit similar behavior, which suggests that iterated predictions from a single model would require a very rich specification (so as to account for the long delays and the switches in sign of when the components become useful for classification) that is likely to be parametrically prohibitive. Instead, the next section investigates the combined predictive ability of the components of the ILI with direct prediction-classification methods.

6.2 Forecasting Business Cycle Turning Points

Let w_t denote the vector of components of the ILI and let $S_t \in \{0, 1\}$ denote the state variables implied by the BCDC-PI dates. In this section we are interested in modeling the posterior probabilities $P[S_{t+h} = s|w_t]$ for $h \in \{0, ..., 24\}$ (we include $h = 0$ as a nowcast). More specifically, we assume the log-odds ratio at time h is a linear function of w_t, so that

$$\log \frac{P[S_{t+h} = 0|w_t]}{P[S_{t+h} = 1|w_t]} = \beta_{h0} + \beta_h' w_t; \qquad h \in \{0, ..., 24\}$$

which results in the well-known logistic model. The parameters of this model can be easily maximized with standard techniques by maximum likelihood or iterated least squares. Moreover, this is a popular model for classification in biostatistics. In fact, linear discriminant analysis (LDA), a standard classification algorithm, consists of the logistic regression we propose and a marginal model for w_t. Hastie, Tibshirani and Friedman (2009) however argue that the logistic model may be a safer choice than LDA. Since most economists are familiar with logistic regression but not necessarily with LDA, we prefer to take the safer route.

The prediction problem over more than one horizon into the future can be done in one of two ways: by specifying the one period ahead model and iterating forward as needed, or by estimating a specific model for each forecast horizon. We prefer to take the latter approach for several reasons. First, the iterative approach would require us to specify a model for w_t that we could use to iterate as well. Second, the specification of the conditional model would have to be sufficiently parametrically intensive to capture the non-monotonicities that we uncovered in the previous section. Third, the nonlinearity of the logistic model would require simulation techniques to construct forecasts beyond one period ahead. This would needlessly complicate the out-of-sample computations we are about to describe. Fourth, as we show in Figure 8, direct-forecasting allows for the signals to be appropriately reweighted depending on the forecast horizon.

The classification-prediction exercise uses a rolling window of fixed width that is used as a training sample. The first window begins January, 1968 and ends December, 1987 (approximately splitting the sample in half since we truncate the sample in November 2007). With this training sample we generate a set of forecasts for $h = 0, ..., 24$ and then roll the training sample by one month and repeat, simultaneously for the two growth transformations we have been investigating all along. We use the collection of out-of-sample classification-predictions to calculate the per-horizon $AUROC$s that are displayed in Figure 9 for both transformations, along with 95% confidence interval bands. The figure shows that the year-on-year growth transformation of the components of the ILI begins with nearly perfect classification ability at $h = 0$ (not surprisingly since in section 3 we discovered that initial claims of unemployment can generate an $AUROC$ of about 0.96), which gradually deteriorates for both transformations as the forecast horizon increases. Over the first year, classification ability remains very high (for the year-on-year transformed components), with $AUROC$s around 0.9. A more steady decline occurs after month 10 or 11 although two years out we still do slightly better than a coin-toss, regardless of transformation. Both transformations behave very similarly after the first year, however. The justification is that classification at the later horizons is mostly driven by the 10-year T-bond-FFR spread, which does not require transformation under either specification.

7 Discussion

The question of how good is the dating of cyclical economic activity proposed by members of the BCDC at the NBER is very difficult to answer: there is no formal definition of what *economic activity* is, and the true state of economic activity (expansion/recession) is unobservable, even retrospectively. This runs counter to the basic principles of statistical evaluation in which predictions from competing models are compared to the realizations observed ex-post in the data and used to determine which specification is best.

This paper offers fresh views on this complex assessment problem from the perspective of evaluating classification ability. The methods that we explore are not intended to guide what econometric methodology should be used to calculate the latent state of economic activity as doing so would require specification of a user-specific loss function. Rather, they are designed to evaluate the outcome of such efforts by providing insights into the classification ability of alternative variables.

Many of the methods used here originate from the biostatistics literature but we make a few methodological contributions of our own. Although we do not propose an operational definition of *economic* *activity* (just a working approximation in the form of a factor model), we are able to formally examine the properties of the sorting of economic indicators implied by the BCDC's chronology (and alternatively cyclical classifiers). This goes a long way toward providing a metric of classification quality and this is novel.

Conditional on the BCDC's chronology (so that the question of what is meant by economic activity and what the true latent state of economic activity is are momentarily set aside), we then provided methods to evaluate the information content of several indices for the purposes of nowcasting and forecasting turning points. Our results are important: they suggest that no single composite index constructed from the components of the ILI will be appropriate for predictions at all horizons. Rather, each horizon requires a different combination of information.

Another contribution of our paper is to organize how one ought to think about the signal obtained from an economic indicator. By considering the underlying decision problem and the relative trade-offs of true and false positives (and implicitly true and false negatives), it is made clear that no single critical value is appropriate for all economic agents. Agents facing different preferences and constraints will make different decisions from the same reading of an index.

We conclude by noting that classification ability does not directly equate with data fit in the manner that is commonly used in economics. A model that has poor fit can nevertheless generate good classification. For this reason, the manner in which information from different variables is combined for the purposes of classification is not necessarily optimized by minimizing a measure of Euclidean distance, such as in principal component analysis or factor models. In current work in progress (Hsieh, Chen, Berge and Jordà, 2010), we are working on decoding algorithms to clarify the differences and offer alternative methods.

8 Appendix

8.1 Data Sources and Calculations

This is a summary of the economic indicators, transformations and data sources provided in the appendix of the December 11, 2008 press release of the Business Cycle Dating Committee of the National Bureau of Economic Analysis and available from their website (www.nber.org).

Indicator	*Sample Available*	*Source and Method*
Industrial Production	1919:1 - 2009:10	FRB index B50001
Real Personal Income less transfers	1959:1 - 2009:5	BEA Table 2.6, line 1 less line 14, both deflated by a monthly interpolation (see below) of BEA Table 1.1.9 line 1
Payroll Employment	1939:1 - 2009:10	BLS Series CES0000000001 (September and October 2009 preliminary)
Household Employment	1948:1 - 2009:10	BLS Series LNS12000000
Real Manufacturing and Trade Sales	1997:1 - 2009:9	BEA Table 2BU, line 1
Real Gross Domestic Product	1947:I - 2009:III	BEA Table 1.1.6, line 1 (2009:III advance estimate)
Real Gross Domestic Income	1947:I - 2009:III	BEA Table 1.10, line 1, divided by BEA Table 1.1.9, line 1 (2009:III advance estimate)

Websites:

- Federal Reserve Board industrial production index:

  `www.federalreserve.gov/releases/g17/iphist/iphist\_sa.txt`

- Bureau of Economic Analysis, U.S. Department of Commerce, all but sales:

  `www.bea.gov/national/nipaweb/SelectTable.asp?Selected=N`

  sales: `www.bea.gov/national/nipaweb/nipa\_underlying/SelectTable.asp`

- BLS payroll survey: `http://data.bls.gov/cgi-bin/surveymost?ce`

- BLS household survey: `http://data.bls.gov/cgi-bin/surveymost?ln`

Interpolation of GDP deflator:

The value of the index in the first month of the quarter is one third of the past quarter's value plus two-thirds of the current quarter's value. In the second month, it is the quarter's value. In the third month, it is two-thirds of the quarter's value plus one third of the next quarter's value.

Indices

Indicator	*Sample Available*	*Source and Method*
Chauvet-Hamilton Index	1967:11 - 2009:2	Chauvet and Hamilton (2005)
Chauvet-Piger Index	1967:2 - 2009:9	Chauvet and Piger (2008)
Aruba Diebold Scotti Index	1960:2 - 2009:10	Federal Reserve Bank of Philadelphia
Chicago Fed National Activity Index	1967:3 - 2009:10	Federal Reserve Bank of Chicago
Purchasing Managers Index	1948:1 - 2009:10	Institute for Supply Management

Websites:

- Chauvet-Hamilton Index: `http://www.econbrowser.com/archives/rec_ind/description.html`
- Chauvet-Piger Index: `http://www.uoregon.edu/~jpiger/us_recession_probs.htm`
- ADS Index: `http://www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/`
- Chicago Fed Index: `http://www.chicagofed.org/economic_research_and_data/cfnai.cfm`
- Purchasing Managers Index: `http://www.ism.ws/`

Conference Board Index of Leading Indicators

Indicator	*Sample Available*
Average weekly hours, manufacturing	1939:1 - 2009:6
Average weekly initial claims for unemployment insurance	1967:1 - 2009:6
Building permits, new private housing units	1960:1 - 2009:6
Index of supplier deliveries—vendor performance	1948:1 - 2009:6
Interest rate spread, 10-year Treasury bonds less federal funds rate	1954:8 - 2009:6
Manufacturer's new orders, consumer goods and materials	1959:1 - 2009:6
Manufacturer's new orders, nondefense capital goods	1959:1 - 2009:6
Money supply, M2	1959:1 - 2009:6
Stock prices, S&P 500	1921:1 - 2009:6
University of Michigan index of consumer expectations	1959:11 - 2009:6

The LexisNexis News Index:

The index is a standardized count of the number of news items that appear in the LexisNexis Academic database (see `http://www.lexisnexis.com/us/lnacademic`). In particular, the count is the number of news articles or news abstracts that LexisNexis retrieves when searching for the word "recession" within "US Newspapers and Wires" source. Our database is at a monthly frequency, beginning in July 1970 and running through June 2009. Each monthly observation is the average daily count for all days within that month, which we then standardize by removing a time trend and adjusting for seasonal variation in the number of counts.

The Principal Components Index: The Principal Components Index is constructed by performing principal components decomposition on the covariance matrix associated with the seven indicators explicitly considered by the BCDC. The sample period for the Principal Components Index is February 1967-July 2009 for the month-on-month index. The sample begins in January 1968 for the index based on the annual growth rates of the indicators, with the different start date due to the difference in the number of log-differences taken when smoothing the data.

References

Aruoba, S. Borağan, Francis X. Diebold and Chiara Scotti (2009) "Real-Time Measurement of Business Conditions," *Journal of Business and Economic Statistics,* 27(4): 417-427.

Baker, Stuart G. and Barnett S. Kramer (2007) "Peirce, Youden, and Receiver Operating Characteristic Curves," *The American Statistician,* 61(4): 343-346.

Bamber, D. (1975) "The Area Above the Ordinal Dominance Graph and the Area Below the Receiver Operating Characteristic Graph," *Journal of Mathematical Psychology,* 12: 387-415.

Baxter, Marianne and Robert G. King (1999) "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *Review of Economics and Statistics,* 81: 575-593.

Burns, Arthur F. and Wesley C. Mitchell (1946) **Measuring Business Cycles.** NBER Book Series Studies in Business Cycles N. 2. New York: NBER.

Business Cycle Dating Committee of the National Bureau of Economic Research, December 11, 2008 press release available at: http://www.nber.org/cycles/dec2008.html.

Chauvet, Marcelle and James D. Hamilton (2005) “Dating Business Cycle Turning Points,” *National Bureau of Economic Research,* working paper 11422.

Chauvet, Marcelle and Jeremy Piger (2008) “A Comparison of the Real-Time Performance of Business Cycle Dating Methods,” *Journal of Business and Economic Statistics,* 26: 42-49.

Diebold, Francis X. and Glenn D. Rudebusch (1990) “A Nonparametric Investigation of Duration Dependence in the American Business Cycle,” *Journal of Political Economy*, 98(3): 596-616.

Elliott, Graham and Robert P. Lieli (2009) “Predicting Binary Outcomes,” U.C.S.D. mimeograph.

Farmer, Roger E. A., Daniel F. Waggoner and Tao Zha (2009) “Understanding Markov-Switching Rational Expectation Models,” *Journal of Economic Theory*, 144(5): 1849-1867.

Green, David M. and John A. Swets (1966) **Signal Detection Theory and Psychophysics.** Peninsula Publishing: Los Altos, California, USA.

Greiner, Matthias, Dirk Pfeiffer and Ronald D. Smith (2000) “Principles and Practical Application of the Receiver Operating Characteristic Analysis for Diagnostic Tests,” *Preventive Veterinary Medicine,* 45: 23-41.

Hamilton, James D. (1989) “A New Approach to the Economic Analysis of Nonstationary Time Series Subject to Changes in Regime,” *Econometrica,* 57(2): 357-384.

Hand, David J. and Veronica Vinciotti (2003) “Local versus Global Models for Classification Problems: Fitting Models Where It Matters,” *The American Statistician*, 57(2): 124-131.

Hanley, James A. and Barbara J. McNeil (1982) “The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve,” *Radiology,* 143: 29-36.

Hastie, Trevor, Robert Tibshirani and Jerome Friedman (2009) **The Elements of Statistical Learning.** Springer: New York, New York, USA.

Hodrick, Robert and Edward C. Prescott (1997) “Postwar U.S. Business Cycles: An Empirical Investigation,” *Journal of Money, Credit and Banking, 29*: 1-16.

Hsieh, Fushing and Bruce W. Turnbull (1996) “Nonparametric and Semiparametric Estimation of the Receiver Operating Characteristic Curve,” *Annals of Statistics*, 24: 25-40.

Hsieh, Fushing, Shu-Chun Chen, Travis J. Berge and Òscar Jordà (2010) “Decoding the Global Business Cycle,” U.C. Davis, mimeograph.

Jordà, Òscar and Alan M. Taylor (2010a) “The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself,” U.C. Davis, NBER working paper 15518.

Jordà, Òscar and Alan M. Taylor (2010b) “Investment Performance of Directional Trading Strategies,” U.C. Davis, *mimeo.*

Khandani, Amir E., Adlar J. Kim and Andrew W. Lo (2010) “Consumer Credit Risk Models via Machine Learning Algorithms,” M.I.T. mimeograph.

Kolmogorov, Andrey N. (1933) “Sulla Determinazione Empirica di una Legge di Distribuzione,” *Giornale dell’Istituto Italiano degli Attuari 4: 83-91.*

Liu, Zheng, Daniel F. Wagonner and Tao Zha (2010) "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach," Emory University, working paper 10-02.

Lusted, Lee B. (1960) "Logical Analysis in Roentgen Diagnosis," *Radiology*, 74: 178-93.

Mann, H. B. and D. R. Whitney (1947) "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," *Annals of Mathematical Statistics*, 18: 50-60.

Mason, Ian B. (1982) "A Model for the Assessment of Weather Forecasts," *Australian Meteorological Society*, 30: 291-303.

Neftçi, Salih N. (1982) "Optimal Prediction of Cyclical Downturns," *Journal of Economic Dynamics and Control*, 4: 225-241.

Obuchowski, Nancy A. (1994) "Computing Sample Size for Receiver Operating Characteristic Curve Studies," *Investigative Radiology,* 29(2): 238-243.

Obuchowski, Nancy A. and Michael L. Lieber (1998) "Confidence Intervals for the Receiver Operating Characteristic Area in Studies with Small Samples," *Academic Radiology,* 5(8): 561-571.

Peirce, Charles S. (1884) "The Numerical Measure of the Success of Predictions," *Science*, 4, 428-441.

Pepe, Margaret S. (2003) **The Statistical Evaluation of Medical Tests for Classification and Prediction.** Oxford, U.K.: Oxford University Press.

Peterson W. Wesley and Theodore G. Birdsall (1953) "The Theory of Signal Detectability: Part I. The General Theory," Electronic Defense Group, Technical Report 13, June 1953. Available from EECS Systems Office, University of Michigan.

Rudebusch, Glenn D. and John C. Williams (2009) "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve," *Journal of Business and Economic Statistics,* 27(4): 492-503.

Shimer, Robert (2005) "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review,* 95(1): 25-49.

Smirnov, Nikolai V. (1939) "Estimate of Deviation Between Empirical Distribution Functions in Two Independent Samples (in Russian)," *Bulletin Moscow University,* 2: 3-16.

Spackman, Kent A. (1989) "Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning," Proceedings of the Sixth International Workshop on Machine Learning. 160-163.

Stanski, Henry R., Laurence J. Wilson and William R. Burrows (1989) "Survey of Common Verification Methods in Meteorology," Research Report No. 89-5, Atmospheric Environment Service, Forecast Research Division, 4905 Dufferin Street, Downsview, Ontario, Canada.

Stock, James H. and Mark W. Watson (1989) "New Indexes of Coincident and Leading Indicators," in **NBER Macroeconomics Annual**, 4: 351-393.

Stock, James H. and Mark W. Watson (1999) "Forecasting Inflation," *Journal of Monetary Economics, 44: 293-335.*

Wilcoxon, Frank (1945) "Individual Comparisons by Ranking Methods," *Biometrics Bulletin,* 1, 80-83.

World Meteorological Organization (2000) **Standardized Verification System (SVS) for Long-Range Forecasts (LRF).** World Meteorological Organization, Geneva, Switzerland.

Wright, Jonathan (2006) "The Yield Curve and Predicting Recessions," Finance and Economics Discussion Series 2006-07*, Federal Reserve Board.*

Youden, William J. (1950) "Index for Rating Diagnostic Tests," *Cancer*, 3: 32-35.

Table 1. AUROCs for Cyclical GDP Using Five Detrending Methods: Annualized Montly Growth, Annual Growth, HP filtered Trend, Baxter-King filtered Trend, and Congressional Budget Office Estimates of Potential Output.

BCDC-PI Chronology					
	Monthly Growth	**Yearly Growth**	**HP Filter**	**BK Filter**	**CBO**
AUROC	0.89	0.91	0.69	0.67	0.74
Std. Error	(0.017)	(0.016)	(0.027)	(0.028)	(0.027)

Max. AUROC by Allowing a Phase Shift of h^* Periods with respect to the BCDC-PI Chronology					
AUROC	0.89	0.98	0.90	0.90	0.85
Std. Error	(0.017)	(0.004)	(0.014)	(0.014)	(0.022)
h^*	0	3	6	7	6
N. Obs	729	718	730	679	706

Notes: Sample period February 1947 to November 2007, except for Potential GDP from the CBO, for which data begins in February 1949. GDP data filtered at quarterly frequency, then interpolated to monthly observations according to BCDC release of December 1, 2008. HP filter calculated with smoothing parameter $\lambda = 1600$. Baxter and King filter set to select frequencies between 6 and 32 quarters. The top panel calculates the AUROC when the chronology of turning points is matched to the BCDC-Peak Included (BCDC-PI) dates. The bottom panel calculates the optimal phase shift (h^*) with respect to the BCDC-PI that maximizes the AUROC in a window of plus or minus 24 months.

Table 2. Four Definitions of Recession. Summary Statistics

	BCDC-PI	**BCDC-PE**	**R1**	**R2**
Number of Recessions	10	10	24	9
Total Months	114	104	108	63
Average length of recession (in months)	11.4	10.4	4.5	7.0

Notes: Sample period February 1947 to November 2007. BCDC-PI refers to NBER recessions defined when the peak and trough months are included (this is the actual definition provided by the BCDC in the December 1, 2008 press release). BCDC-PE refers to NBER recessions where the peak date is excluded. R1 refers to the mechanical rule that classifies a recession as any observation where GDP growth is negative. R2 is the rule that instead requires two consecutive quarters of negative growth.

Table 3. Four Definitions of Recession: AUROCs

No Phase Shift: *h = 0*	**Month-on-month**				**Year-on-year**			
	BCDC-PI	**BCDC-PE**	**R1**	**R2**	**BCDC-PI**	**BCDC-PE**	**R1**	**R2**
GDP	0.89	0.90	0.89	0.91	0.91	0.93	0.83	0.90
(s.e.)	(0.017)	(0.017)	(0.021)	(0.022)	(0.016)	(0.013)	(0.025)	(0.022)
GDI	0.92	0.93	0.87	0.90	0.90	0.92	0.80	0.89
(s.e.)	(0.015)	(0.013)	(0.022)	(0.022)	(0.017)	(0.014)	(0.027)	(0.021)
PI	0.83	0.86	0.81	0.84	0.83	0.85	0.76	0.84
(s.e.)	(0.026)	(0.024)	(0.030)	(0.033)	(0.021)	(0.019)	(0.031)	(0.027)
IP	0.85	0.87	0.79	0.83	0.86	0.89	0.72	0.77
(s.e.)	(0.020)	(0.019)	(0.028)	(0.031)	(0.019)	(0.017)	(0.031)	(0.037)
MTS	0.74	0.75	0.70	0.75	0.92	0.94	0.82	0.91
(s.e.)	(0.034)	(0.035)	(0.042)	(0.051)	(0.016)	(0.012)	(0.035)	(0.021)
PE	0.90	0.92	0.78	0.84	0.80	0.82	0.66	0.71
(s.e.)	(0.017)	(0.014)	(0.030)	(0.032)	(0.021)	(0.019)	(0.031)	(0.035)
HE	0.73	0.74	0.69	0.77	0.76	0.78	0.66	0.77
(s.e.)	(0.027)	(0.027)	(0.030)	(0.032)	(0.025)	(0.025)	(0.033)	(0.032)
PCI	0.94	0.96	0.89	0.93	0.87	0.89	0.78	0.87
(s.e)	(0.020)	(0.014)	(0.029)	(0.027)	(0.022)	(0.019)	(0.037)	(0.024)
Phase Shift Allowed: h^*								
	BCDC-PI	**BCDC-PE**	**R1**	**R2**	**BCDC-PI**	**BCDC-PE**	**R1**	**R2**
GDP	0.89	0.90	1.00	0.98	0.98	0.99	0.87	0.98
(s.e.)	(0.017)	(0.017)	(0.000)	(0.005)	(0.004)	(0.004)	(0.022)	(0.005)
h^*	0	0	-1	-1	3	3	4	4
GDI	0.92	0.93	0.95	0.95	0.98	0.98	0.86	0.97
(s.e.)	(0.015)	(0.013)	(0.010)	(0.009)	(0.005)	(0.005)	(0.023)	(0.007)
h^*	0	0	-1	-1	4	3	5	4
PI	0.84	0.86	0.81	0.84	0.96	0.96	0.85	0.93
(s.e.)	(0.024)	(0.024)	(0.030)	(0.033)	(0.009)	(0.008)	(0.028)	(0.015)
h^*	1	0	0	0	6	6	6	6
IP	0.85	0.87	0.82	0.84	0.96	0.97	0.83	0.91
(s.e.)	(0.020)	(0.019)	(0.024)	(0.028)	(0.008)	(0.008)	(0.024)	(0.021)
h^*	0	0	-1	-1	4	4	6	6
MTS	0.74	0.75	0.74	0.76	0.96	0.96	0.88	0.95
(s.e.)	(0.034)	(0.035)	(0.037)	(0.047)	(0.009)	(0.010)	(0.026)	(0.010)
h^*	0	0	-1	-2	3	2	5	4
PE	0.90	0.92	0.78	0.85	0.96	0.96	0.82	0.90
(s.e.)	(0.018)	(0.014)	(0.030)	(0.028)	(0.008)	(0.008)	(0.025)	(0.022)
h^*	1	0	0	-1	6	6	6	6
HE	0.73	0.74	0.69	0.77	0.94	0.94	0.82	0.77
(s.e.)	(0.028)	(0.027)	(0.030)	(0.032)	(0.014)	(0.014)	(0.027)	(0.032)
h^*	2	0	1	0	6	6	6	6
PCI	0.94	0.96	0.95	0.97	0.98	0.98	0.89	0.96
(s.e.)	(0.020)	(0.014)	(0.017)	(0.009)	(0.005)	(0.005)	(0.029)	(0.011)
h^*	0	0	-1	-1	6	5	6	6

Notes: Sample period February 1947 to November 2007. BCDC-PI refers to NBER recessions defined when the peak and trough months are included (this is the actual definition provided by the BCDC in the December 1, 2008 press release). BCDC-PE refers to NBER recessions where the peak date is excluded. R1 refers to the mechanical rule that classifies a recession as any observation where GDP growth is negative. R2 is the rule that instead requires two consecutive quarters of negative growth. Top panel does not allow for phase shift with respect to the recession dates associated with each rule. The bottom panel reports the phase shift associated with the maximum AUROC by looking into a window of plus or minus 24 months. AUROC values with standard errors in parenthesis and in the bottom panel we also report the value of the phase shift h^*.

Table 4. Comparing the BCDC to Hidden-Markov Mixture Model Turning Point Predictions

No Phase Shift: $h = 0$						
		Month-on-month			**Year-on-year**	
	BCDC-PI	**CH**	**CP**	**BCDC-PI**	**CH**	**CP**
GDP	0.89	0.89	0.75	0.91	0.94	0.81
(s.e.)	(0.017)	(0.021)	(0.022)	(0.016)	(0.015)	(0.019)
GDI	0.92	0.91	0.81	0.90	0.92	0.82
(s.e.)	(0.015)	(0.021)	(0.019)	(0.017)	(0.017)	(0.019)
PI	0.83	0.81	0.71	0.83	0.83	0.74
(s.e.)	(0.026)	(0.027)	(0.023)	(0.021)	(0.023)	(0.022)
IP	0.85	0.82	0.74	0.86	0.85	0.83
(s.e.)	(0.020)	(0.030)	(0.022)	(0.019)	(0.024)	(0.019)
MTS	0.74	0.73	0.66	0.92	0.91	0.86
(s.e.)	(0.034)	(0.033)	(0.025)	(0.016)	(0.015)	(0.017)
PE	0.90	0.88	0.80	0.80	0.76	0.73
(s.e.)	(0.017)	(0.021)	(0.020)	(0.021)	(0.029)	(0.022)
HE	0.73	0.74	0.65	0.76	0.75	0.70
(s.e.)	(0.027)	(0.032)	(0.025)	(0.025)	(0.034)	(0.023)
PCI	0.94	0.96	0.96	0.87	0.88	0.90
(s.e.)	(0.020)	(0.012)	(0.013)	(0.022)	(0.021)	(0.016)
Phase Shift Allowed: h^*						
	BCDC-PI	**CH**	**CP**	**BCDC-PI**	**CH**	**CP**
GDP	0.89	0.89	0.75	0.98	0.98	0.89
(s.e.)	(0.017)	(0.021)	(0.022)	(0.004)	(0.006)	(0.014)
h^*	0	0	-1	3	3	6
GDI	0.92	0.91	0.81	0.98	0.97	0.93
(s.e.)	(0.015)	(0.021)	(0.019)	(0.005)	(0.007)	(0.010)
h^*	0	0	1	4	3	6
PI	0.84	0.82	0.73	0.96	0.94	0.89
(s.e.)	(0.024)	(0.027)	(0.023)	(0.009)	(0.011)	(0.015)
h^*	1	1	3	6	5	6
IP	0.85	0.82	0.74	0.96	0.95	0.95
(s.e.)	(0.020)	(0.031)	(0.022)	(0.008)	(0.012)	(0.009)
h^*	0	1	0	4	4	6
MTS	0.74	0.73	0.66	0.96	0.94	0.93
(s.e.)	(0.034)	(0.033)	(0.025)	(0.009)	(0.012)	(0.011)
h^*	0	0	0	3	2	5
PE	0.90	0.90	0.84	0.96	0.94	0.90
(s.e.)	(0.018)	(0.021)	(0.018)	(0.008)	(0.013)	(0.013)
h^*	1	1	5	6	6	6
HE	0.73	0.75	0.68	0.94	0.93	0.84
(s.e.)	(0.028)	(0.032)	(0.024)	(0.014)	(0.017)	(0.017)
h^*	2	1	4	6	6	6
PCI	0.94	0.96	0.96	0.98	0.97	0.98
(s.e.)	(0.020)	(0.012)	(0.013)	(0.005)	(0.007)	(0.006)
h^*	0	0	0	6	5	4

Notes: Refer to the appendix for the sample period of each individual series. BCDC-PI refers to NBER recessions defined when the peak and trough months are included (this is the actual definition provided by the BCDC in the December 11, 2008 press release). CH refers to Chauvet and Hamilton's (2005) business cycle dates. CP refers to Chauvet and Piger's (2008) business cycle dates. Top panel does not allow for phase shift with respect to the recession dates associated with each rule. The bottom panel reports the phase shift associated with the maximum AUROC by looking into a window of plus or minus 24 months. AUROC values with standard errors in parenthesis and in the bottom panel we also report the value of the phase shift h^*.

Table 5. Classification Ability of Current Business Conditions Indices: Chicago Fed National Activity Index (CFNAI), Aruoba, Diebold and Scotti (ADS) Index, Purchasing Managers Index (PMI), and LexisNexis Index

Areas Under the ROC Curve

	CFNAI	ADS	PMI	LexisNexis
h = 0	0.93	0.96	0.90	0.78
	(0.02)	(0.01)	(0.02)	(0.03)
h = 12	0.48	0.56	0.48	0.42
	(0.04)	(0.04)	(0.03)	(0.04)

Notes: CFNAI refers to the CFNAI-MA3 version of the index (see appendix for the source). Standard errors reported in parenthesis. See appendix for specific sample period of each indicator.

Classification since December 2007 (last release from BCDC) until October 2009

	ADS	CFNAI	PMI
Threshold	-0.80	-0.72	44.5
December-07	-0.20	-0.44	49.1
January-08	-0.56	-0.40	50.8
February-08	-1.06	-0.82	48.8
March-08	-1.03	-0.99	49.0
April-08	-0.90	-1.12	48.6
May-08	-0.93	-1.12	49.3
June-08	-0.92	-1.07	49.5
July-08	-1.05	-1.16	49.5
August-08	-1.92	-1.41	49.3
September-08	-3.36	-2.24	43.4
October-08	-1.78	-2.31	38.7
November-08	-1.94	-2.63	36.6
December-08	-2.95	-2.74	32.9
January-09	-3.54	-3.63	35.6
February-09	-2.91	-3.46	35.8
March-09	-2.52	-3.32	36.3
April-09	-1.90	-2.66	40.1
May-09	-1.64	-2.64	42.8
June-09	-1.26	-2.13	44.8
July-09	-0.21	-1.53	48.9
August-09	0.33	-0.94	52.9
September-09	0.06	-0.67	52.6
October-09	-0.33	-0.91	55.7

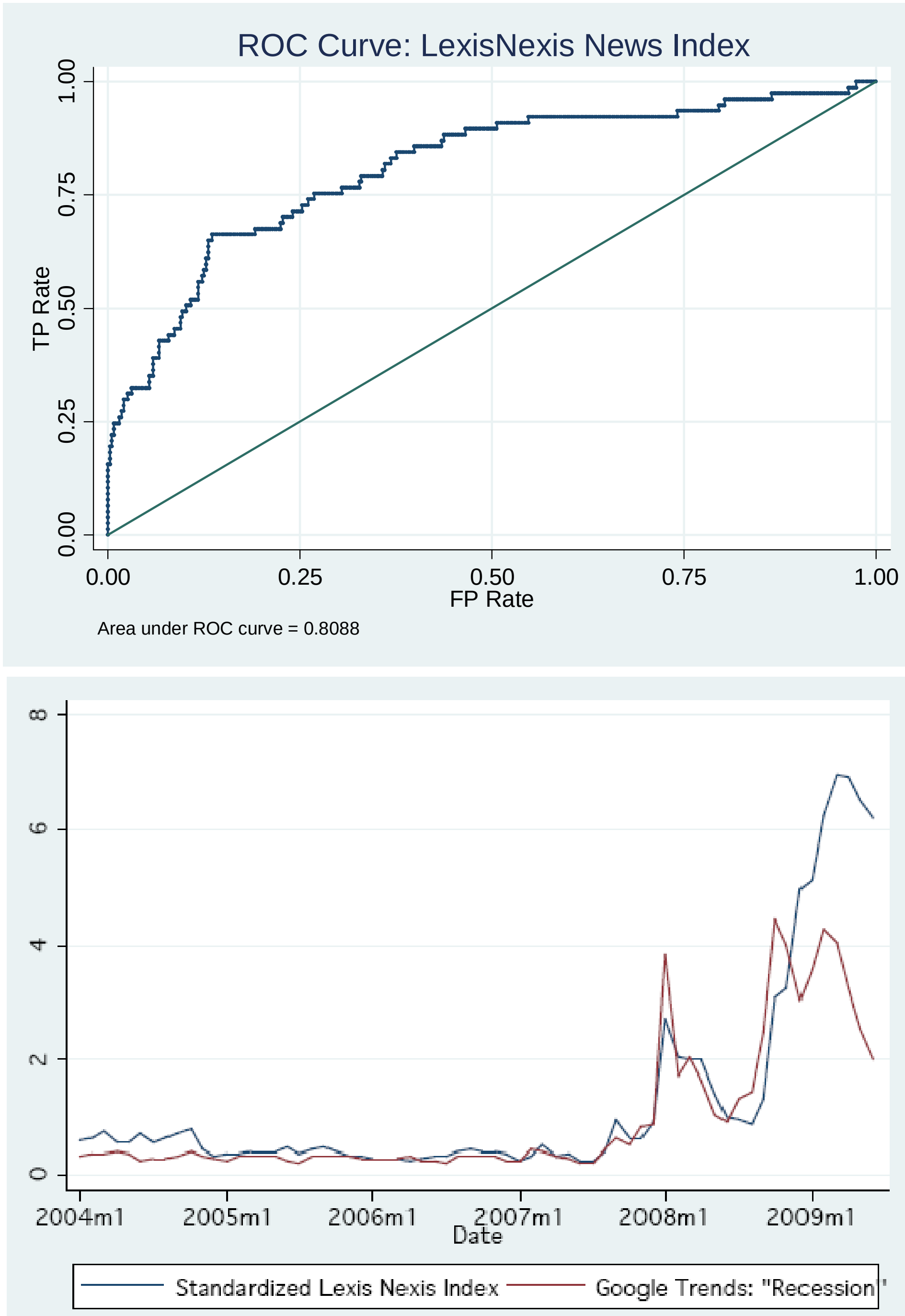
Notes: Threshold refers to the value of the index that maximizes the utility of the method when hits and misses are given the same weight in absolute value. Hence, ADS would suggest we emerged from the recession July 2009, CFNAI September 2009 and PMI June 2009. The shaded dates indicate values of the index below the threshold.

Table 6. Classification Ability of the Components of the Index of Leading Indicators

	Max/Min AUROC			
	Month-on-month		Year-on-year	
	Max	**Min**	**Max**	**Min**
10year T-bond FFR Spread	0.73	0.67	0.73	0.67
(s.e)	(0.02)	(0.03)	(0.02)	(0.03)
horizon	18	0	18	0
Initial Claims of Unemployment	0.70	0.61	0.96	0.71
(s.e)	(0.04)	(0.04)	(0.01)	(0.04)
horizon	0	14	2	24
Consumer Expectations	0.53	0.60	0.82	0.80
(s.e.)	(0.04)	(0.04)	(0.03)	(0.03)
horizon	0	6	0	14
New Private Housing Units	0.53	0.61	0.81	0.80
(s.e.)	(0.04)	(0.04)	(0.03)	(0.03)
horizon	0	7	0	15
Avg. weekly hours, manufacturing	0.66	0.67	0.93	0.84
(s.e.)	(0.03)	(0.03)	(0.01)	(0.02)
horizon	0	9	2	17
M2	0.59	0.66	0.75	0.72
(s.e.)	(0.04)	(0.03)	(0.03)	(0.03)
horizon	0	14	0	22
Vendor Performance	0.55	0.67	0.76	0.79
(s.e.)	(0.03)	(0.03)	(0.03)	(0.03)
horizon	0	9	0	15
New orders: capital goods	0.61	0.55	0.89	0.57
(s.e.)	(0.04)	(0.04)	(0.02)	(0.03)
horizon	1	14	4	22
New orders: consumer goods	0.68	0.59	0.93	0.71
(s.e.)	(0.03)	(0.03)	(0.01)	(0.03)
Horizon	0	12	1	22
S&P 500	0.55	0.62	0.86	0.73
(s.e.)	(0.03)	(0.03)	(0.02)	(0.03)
horizon	0	7	0	14

Notes: All components of the ILI (except the 10-year T-bond – FFR spread) transformed by taking first log difference (month-on-month columns) or the twelfth log difference (year-on-year columns). Standard errors reported in parenthesis. The entry “horizon” refers to the future horizon where the AUROC is maximized/minimized. Minimum AUROCs, if less than 0.5 are reported using the reciprocal in the usual [0.5,1] interval. See appendix for the sample period of each individual series.

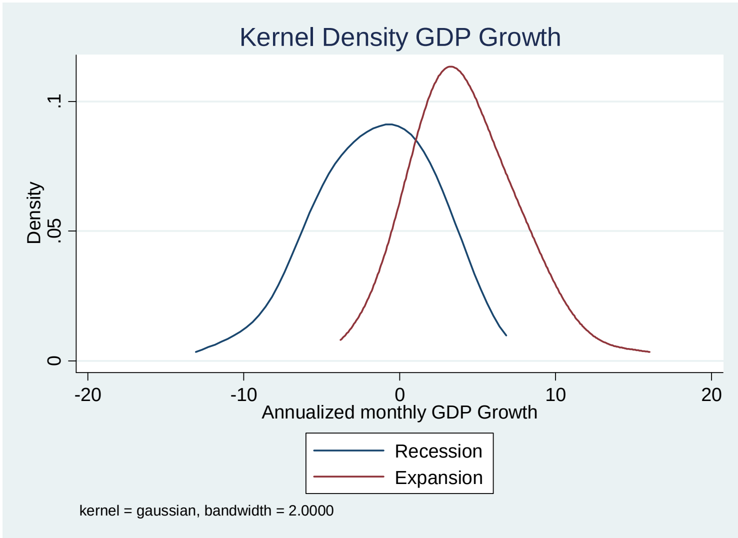
Figure 1. The ROC Curve for the LexisNexis News Index of the Word "Recession"



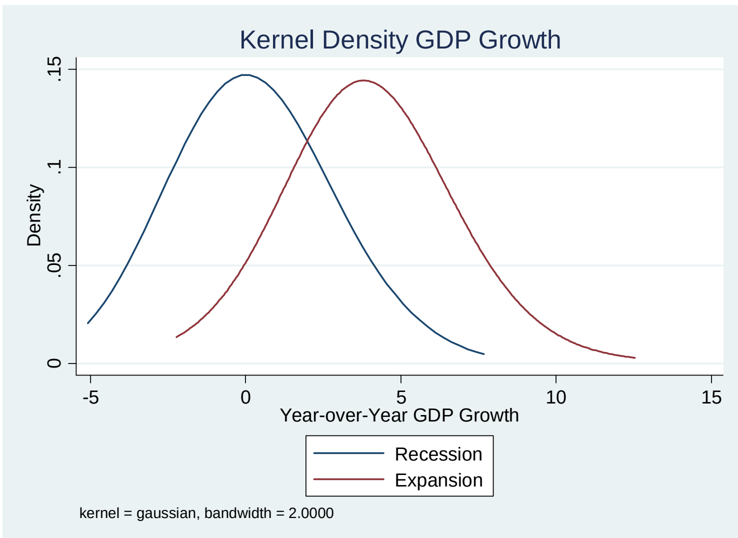
Note: Top panel: ROC curve for the index constructed with the number of news items containing the word "recession" in the LexisNexis academic database (for more details see the appendix). The sample begins July 1970 and ends June 2009. The area under the ROC curve is 0.81. Bottom Panel: plot of the LexisNexis index and the Google Trends index for the word "recession" for the longest sample available from Google Trends (www.google.com/trends).

Figure 2. Kernel Density Estimates of the Mixture Distribution of GDP Growth Implied by Expansion/Recession NBER dates. Sample: March 1947 to November 2007

Annualized Monthly GDP Growth in Percent

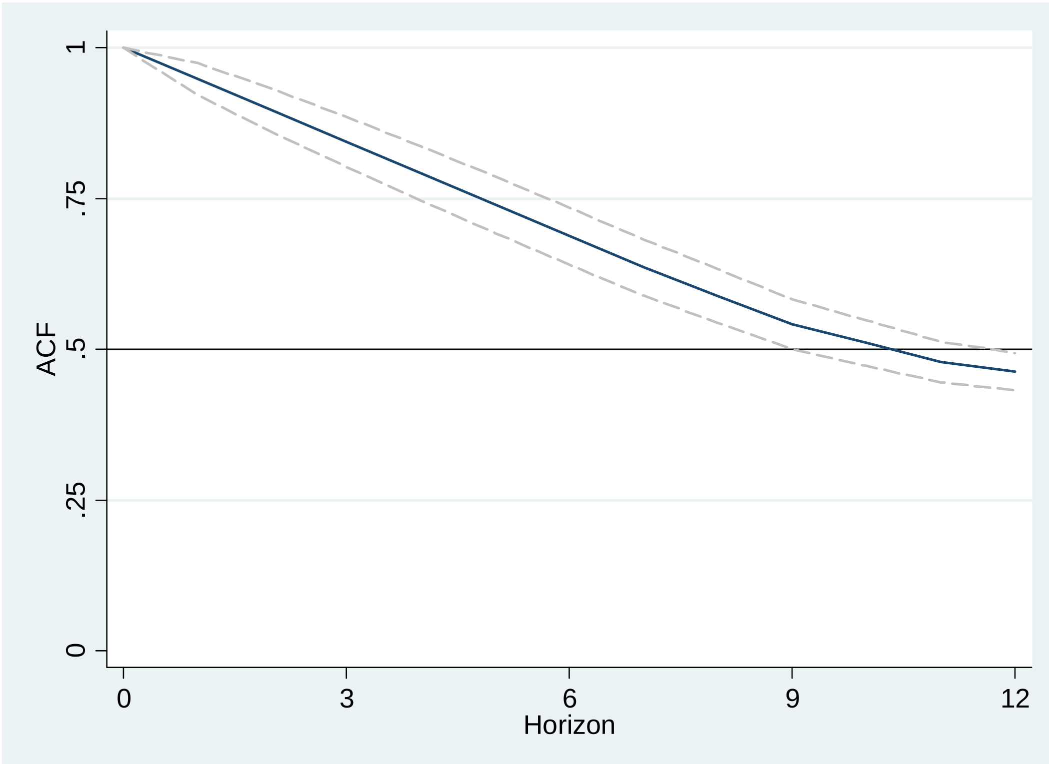


Year-on-year GDP Growth in Percent



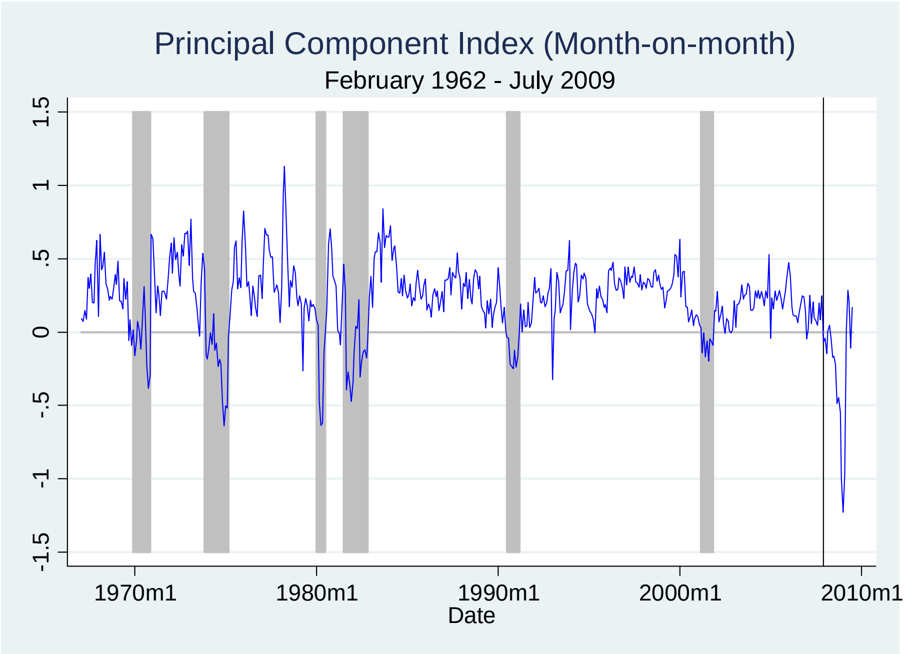
Notes: Monthly GDP growth constructed by interpolating GDP into monthly data using the BCDC linear interpolation method (see December 1, 2008 release). We take the log difference between consecutive months, annualize it and then express it in annual percentages. The recession mean/standard deviation is -1.6/3.55 and the expansion mean/standard deviation is 4.2/3.25. Yearly GDP growth constructed from the monthly interpolated data by taking the twelfth log difference expressed in percentages. The recession mean/standard deviation is 0.05/2.11 and the expansion mean/standard deviation is 3.9/1.97.

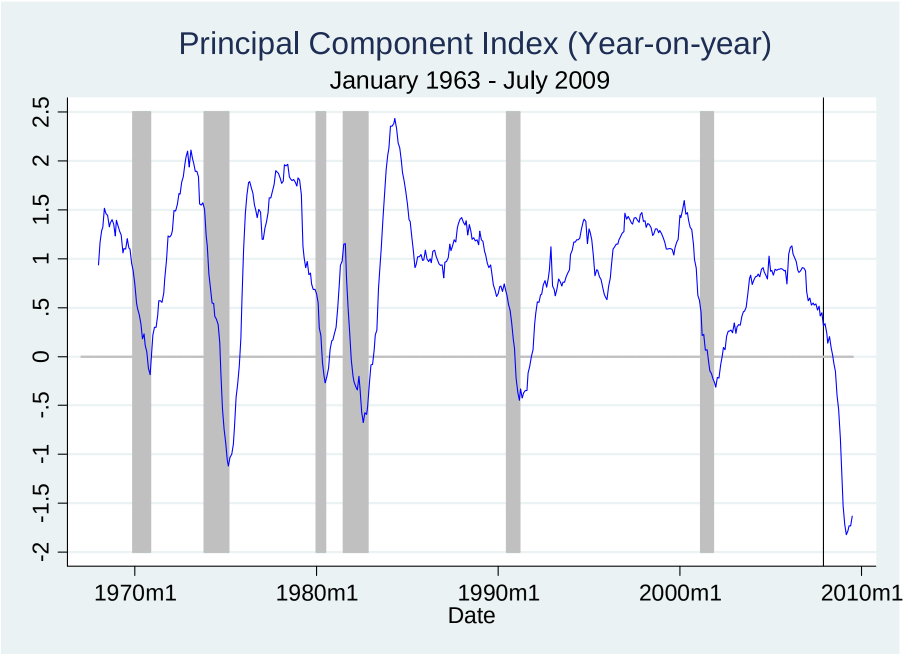
Figure 3. Plot of the AutoClassification Function for BCDC-PI Recession Dates



Notes: The AutoClassification Function (ACF) is a plot of the AUROC of S_t using S_{t-h} for $h > 0$ as the classifier. An AUROC = 0.5 means there is no serial classification ability. A value close to 1 reflects near perfect serial classification ability. The value of the AUROC for h = 1 is 0.95.

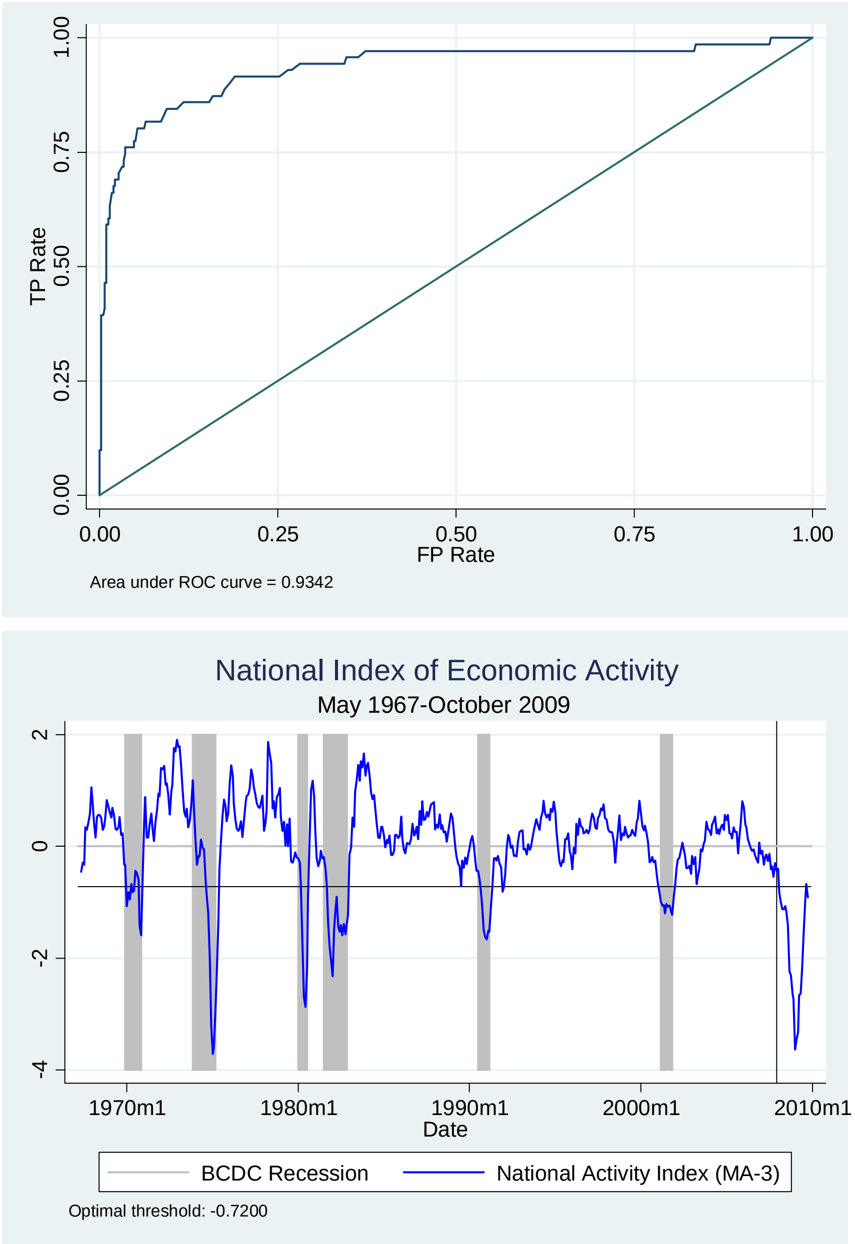
Figure 4. Principal Component Indexes of Economic Activity





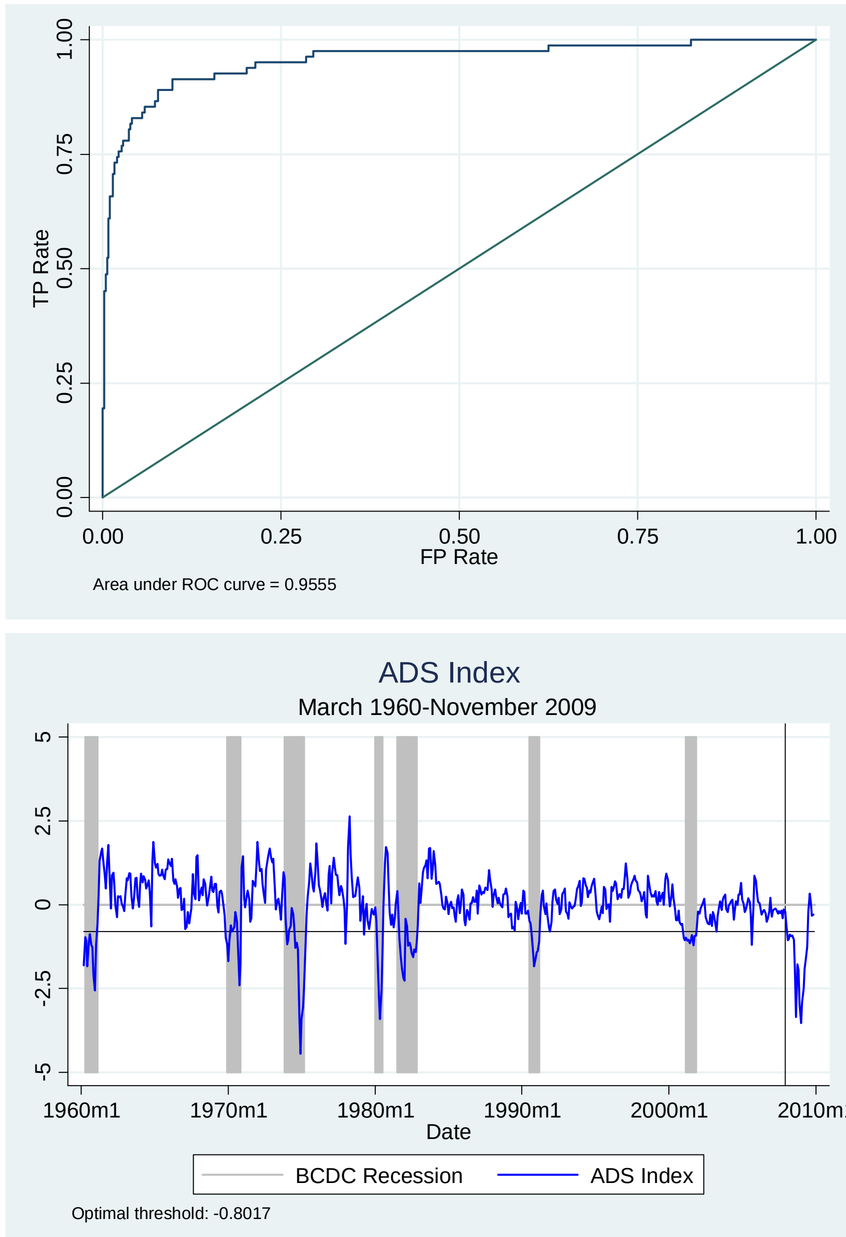
Notes: Principal component indexes are constructed as the first factor from a principal component analysis of the 7 economic indicators used by the BCDC. For more details, consult the appendix.

Figure 5.The Chicago Fed National Activity Index (CFNAI): ROC Curve and Time Series Plot with Optimal Threshold



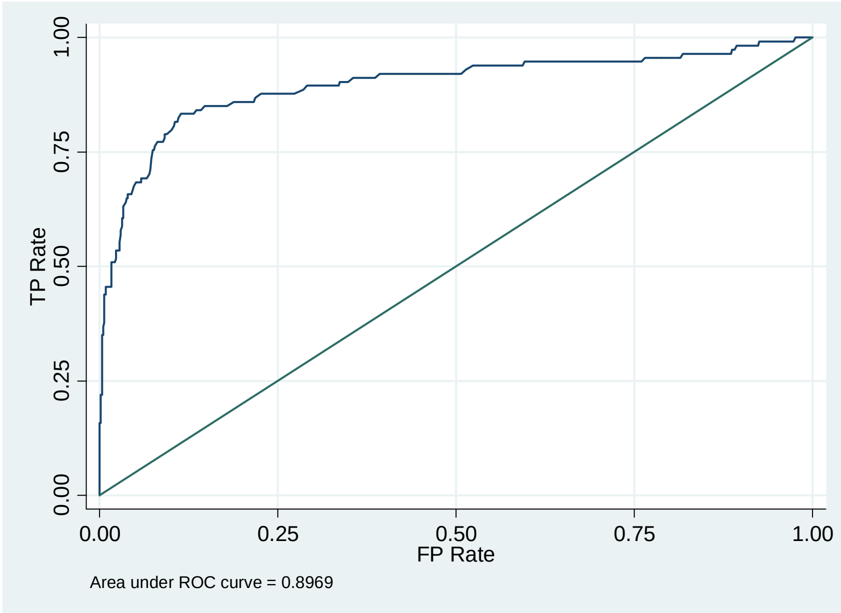
Notes: Top panel: The ROC curve of the BCDC-PI dating using the CFNAI as the classifier. The AUROC = 0.9342. Bottom Panel: Time series plot of the CFNAI. The horizontal line corresponds to the value of the index that maximizes the utility of the method, assuming equally weighted benefits and costs, and calculated with data prior to December 2007 (vertical line). The optimal threshold is -0.72 so that values above it would be classified as expansions, values below as recessions. The first value above this threshold occurs in September 2009.

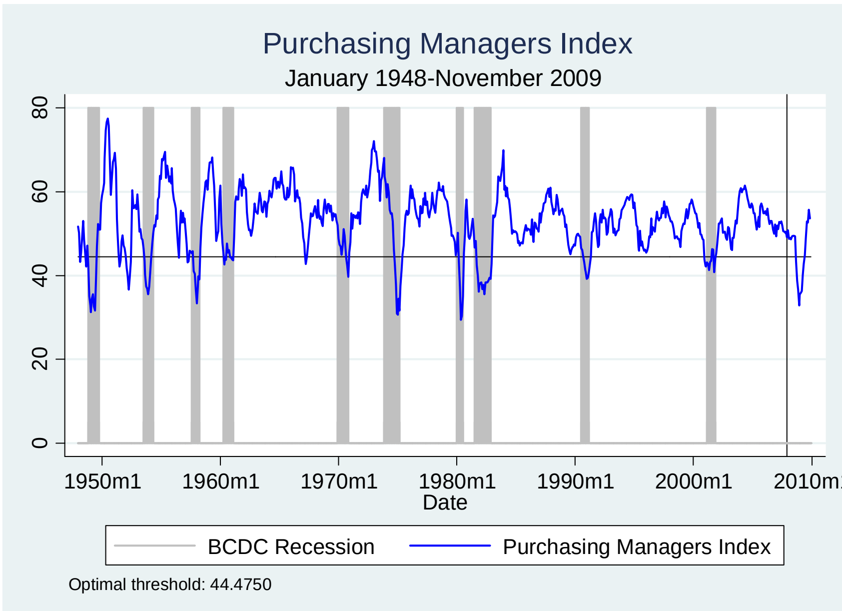
Figure 6. The Aruoba, Diebold and Scotti (ADS) Index: ROC Curve and Time Series Plot with Optimal Threshold



Notes: Top panel: The ROC curve of the BCDC-PI dating using the ADS as the classifier. The AUROC = 0.9555. Bottom Panel: Time series plot of the ADS. The horizontal line corresponds to the value of the index that maximizes the utility of the method, assuming equally weighted benefits and costs and calculated with data prior to December 2007 (vertical line). The optimal threshold is -0.80 so that values above it would be classified as expansions, values below as recessions. The first observation above this threshold occurs in July 2009.

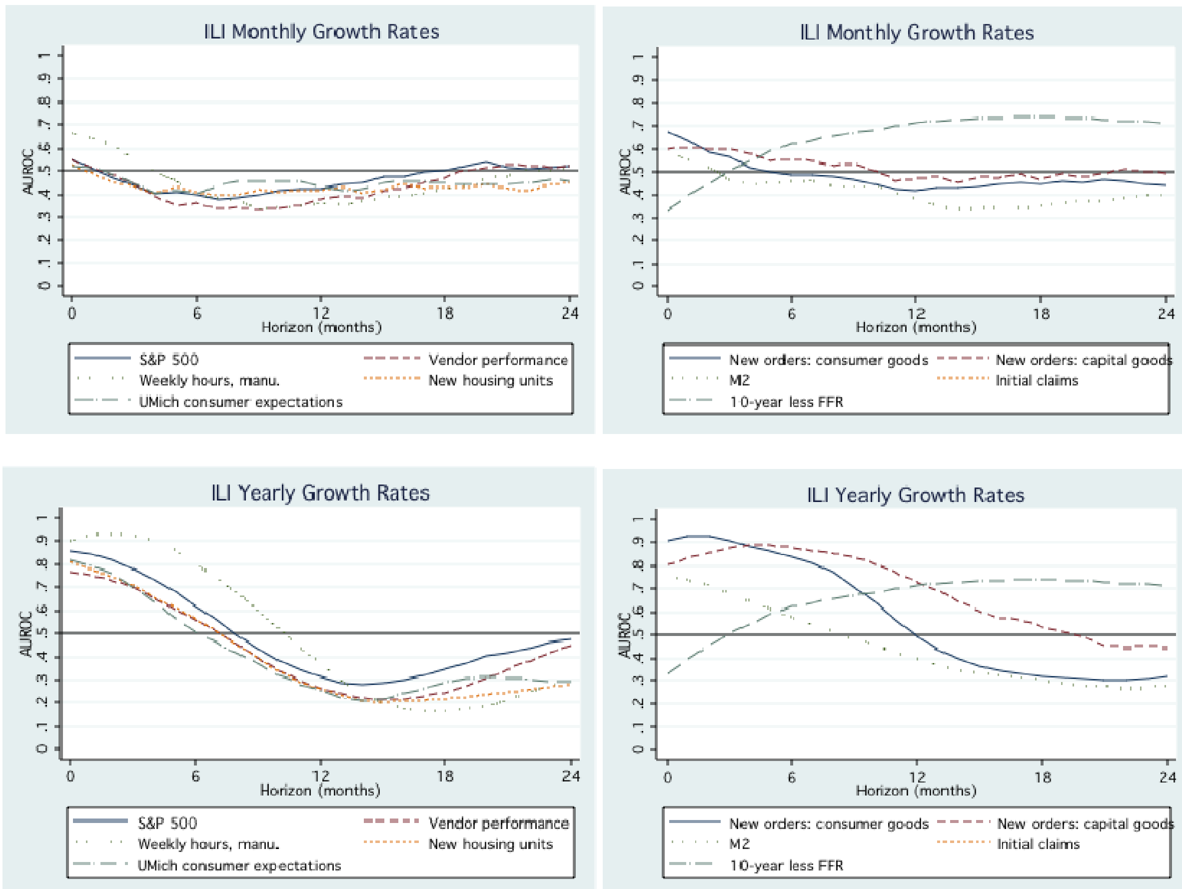
Figure 7. The Purchasing Managers Index (PMI): ROC Curve and Time Series Plot with Optimal Threshold





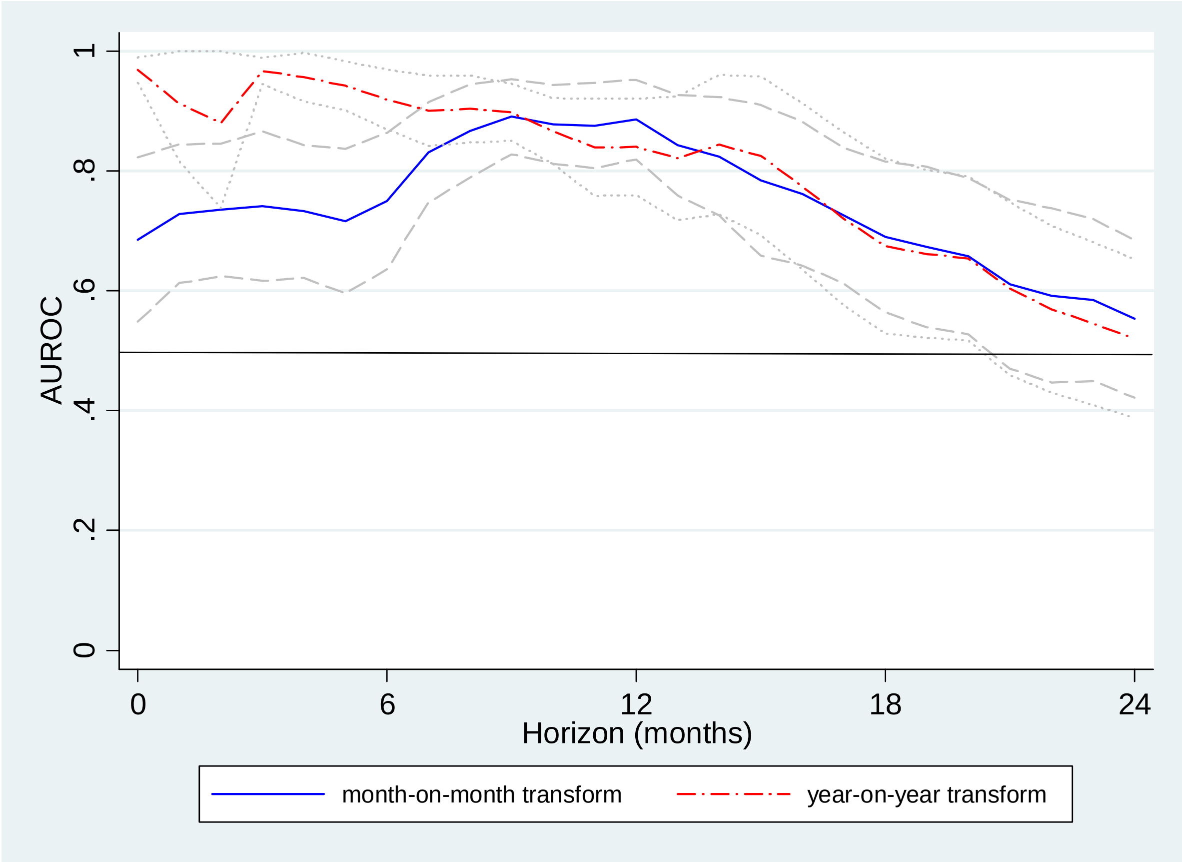
Notes: <u>Top panel</u>: The ROC curve of the BCDC-PI dating using the PMI as the classifier. The AUROC = 0.8969. <u>Bottom Panel:</u> Time series plot of PMI. The horizontal line corresponds to the value of the index that maximizes the utility of the method, assuming equally weighted benefits and costs and calculated with data prior to December 2007 (vertical line). The value of this optimal threshold is 44.48 so that values above it would be classified as expansions, values below as recessions. The first observation above this threshold occurs in June 2009.

Figure 8. AUROC Plots 0 to 24 Months into the Future for the Components of the Index of Leading Indicators



Notes: The top row uses as the classifier the annualized, month-on-month growth rate for each component of the ILI. The bottom panel uses annual growth rates instead. The exception is the 10-year T-Bond – Federal Funds Rate spread, which we do not transform into a growth rate. Each row is broken into two columns, and confidence intervals are suppressed, to facilitate readability. Sample period 1947:2-2007:11.

Figure 9. Out-of-Sample Classification Ability of the Components of the Index of Leading Indicators



Notes: Out-of-sample values of the AUROC and 95% confidence bands 0 to 24 periods into the future with rolling window logistic regressions starting with the sample January 1968 to December 1987 and ending November 2007. Month-on-month transform refers to the direct logistic regressions where the components of the ILI are transformed by taking the log first difference whereas year-on-year transform is done by taking the log twelfth difference.