December 14, 2009

# The Classification of Economic Activity *


**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, Richard Dennis, Graham Elliott, John Fernald, Pierre Olivier Gourinchas, James Hamilton, Fushing Hsieh, Kevin Lansing, Jeremy Piger, Glenn Rudebusch, Alan Taylor, Hal Varian, John Williams and 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 (including 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.

This paper asks three important questions about cyclical economic activity: (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?

We find that dating when recessions begin and end depends critically on how cyclical activity is constructed. Month-to-month variation in economic data chronologically matches the BCDC dating, but it is noisy and hence harder to classify than year-to-year variation. However, the smoother year-to-year variation suggests that the beginning and end of recessions should be translated forward by three months. If instead cycles are derived from detrending methods common in the real business cycle literature, such as the Hodrick-Prescott (1997) and Baxter and King (1999) filters, the timing of recessions should be translated forward by six-months, although separating economic data into expansions and recessions then becomes more difficult than even with the month-on-month analysis. Second, we find that popular indices of business conditions, such as the Aruoba, Diebold and Scotti (ADS) index of business conditions[1] and the Chicago Fed National Activity Index[2] (CFNAI), provide accurate signals about the current state of the business cycle. Finally, we find that the ability to detect future turning points varies across the components of the Conference Board's Index of Leading Indicators (ILI) and changes depending on the forecast horizon considered. At some horizons, positive values of a given index predict when

[1] www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/
[2] www.chicagofed.org/economic_research_and_data/cfnai.cfm

the economy is likely in recession while at other horizons they predict when the economy is likely in expansion instead. These findings suggest that conventional, parsimonious, affine model specifications lack sufficient texture to take full advantage of the predictive information contained in the ILI. We provide out-of-sample evidence about direct predictive-classification ability up to 24 months into the future.

The desire to keep a chronology of turning points – peaks versus troughs of economic activity and hence implicitly the classification of historical economic time series into periods of expansion and recession – reflects the notion that there are fundamental differences between these two phases of the economic cycle. Otherwise the dating of business cycles would amount to a mindless, mechanical, accounting exercise about when GDP growth is observed to be negative. The BCDC's definition of a recession[3] 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.*

This definition, which harkens back to Burns and Mitchell (1946), makes clear that the BCDC does not simply take, for example, a negative observation of industrial production to indicate that the economy is in recession — that same negative datum for industrial production will sometimes be classified as belonging to an expansion and other times as belonging to a recession. It is this classification of economic activity into expansions and recessions that suggests economic activity can be thought of as coming from a mixture of two distinct distributions, a feature that we take advantage of in our analysis. Moreover, information regarding a binomial variable describing aggregate economic activity is simple and easily understood by both policy-makers and the general public: policy-makers may prefer to craft policy responses with a probabilistic statement regarding the recession/expansion state of the economy than with a more uncertain point-estimate of quarterly growth in GDP.

The methods we use in this paper are mostly new to economics, 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

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

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 (2009a, 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 are rely on the specification of an underlying forecast loss function. A major contribution of our paper is to introduce a set of statistical tools based on ROC analysis that offer several advantages over these traditional measures. The ROC curve is independent of the forecast loss function, providing a non-parametric method for judging different potential classification indices. In addition, strictly monotone transformations of the same prediction index have the same ROC curve: these new evaluation methods are not directly tied to modeling ability but to the information content of the indices themselves and automatically encompass a larger class of specifications – the main focus of this paper. 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% error rate (as it misses all the recessions). Our methods are set-up to explicitly recognize the policy trade-offs of these two error rates.

## 2 Classification Ability: The ROC Curve

The methods that we use in this paper will 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 (DGP), 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 and hence, the decision problem is independent of the loss function one may consider. We now explain our approach in detail by discussing first how to evaluate indicators taking the BCDC's dating to be the true classification of business cycles before discussing 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 (albeit with a considerable delay, as we know). Meanwhile, consider the index $Y_t$, which we require only to be real-valued and ordinal. $Y_t$ 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 to 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 $c \rightarrow \infty$, $TP(c) = FP(c) = 0$. Conversely, when $c \rightarrow -\infty$, $TP(c) = FP(c) = 1$, 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. 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 below the $45^0$ diagonal. However in that case it is easy to see that reversing the predictions from the classifier would generate a ROC curve above this 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 news items with the word "recession" appearing in the LexisNexis database every month.[4] 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[5] index for the word recession over the longest sample available for 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 $j$, $i, j \in \{0, 1\}$ and $\pi$ 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

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

[4] 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.

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

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.

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, non-parametric 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):

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

$$\sigma^2 = \frac{1}{n_0 n_1} AUROC(1 - AUROC) + (n_1 - 1)(Q_1 - AUROC^2) + (n_0 - 1)(Q_2 - AUROC^2)$$

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

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, 2009b 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 the final word on the 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.[6] Here we propose a possible solution to this problem.

We begin by taking the view that economic activity can be approximately represented by a 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[7] begins March 1947/February 1948 and ends August 2009 (the difference in start dates reflects the monthly/yearly growth rate

[6] 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."

[7] 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.

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 classification has very high skill.

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 $AUROC = 1$. Because a considerable portion of the paper consists in evaluating potential classifiers of the true state of the economy based on the BCDC dates, we use this metric to assess the skill of 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 Assessing the Business Cycle Dating Committee

There are three characteristics of recessions that will 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; and (3) different definitions of what a recession is have varying classification skill.

The first of these factors can be illustrated with a novel concept that we introduce and that we call "the autoclassification function[8]" (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. A natural benchmark is that if knowing the state in a previous period is not useful to classify the state in 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 for up

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

to 12 months and shows clearly that past information has considerable classification ability about the likelihood of the current state up to about 8 months into the past (at 9 to 12 months the $AUROC$ is virtually 0.5).

This finding has two consequences. First, state-dependence dies-off sufficiently quickly relative to the grouping of recessions and expansions over the entire sample so as to have negligible effect on the regularity conditions[9] in Hsieh and Turnbull (1996) that are required to obtain the basic asymptotic results for the $AUROC$ in expression (5). However, it suggests that past information about the state of the economy is useful for short-run classification. Indeed this is the view taken, for example, in Hamilton's (1989) well-worn hidden Markov mixture model which allows transitions across states to be Markovian. However, the focus of our paper is not on modelling but on evaluation and we felt it was preferable to set this form of time dependence aside so that our results would be least controversial. Even setting this higher hurdle for ourselves, we still find substantial classification ability in the indicators that we investigate.

Therefore, we begin this section by showing that year-on-year growth rate transformations are best for sorting the economy into expansions and recessions, although the timing of the BCDC chronology is best captured by the noisier month-to-month transformation. Other detrending methods do not improve classification ability and result in even greater phase shifts with respect to the BCDC chronology than the year-on-year transformation. Moreover, these alternative detrending methods vary considerably depending on the sample considered and often require information about future values (since they are based on double-sided filters), making them practically undesirable as well.

We next evaluate four alternative definitions of what a recession is before we settle on the definition provided by the December 11, 2008 release of the BCDC: a recession is the period from peak to trough, including the peak and trough months. We do this to maintain comparability with other studies despite finding slightly better classification ability when the definition is modified to exclude peak months. Having settled on the appropriate detrending method (we will simultaneously report month-on-month and year-on-year growth rate transformations from here on) and the definition of recession (peaks and through months included), we then compare the BCDC dating against popular statistical-based hidden-Markov mixture specifications and show that the BCDC does comparably very well across the major economic indicators considered by the BCDC.

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

## 3.1 Trends and Cycles

In a growing economy, classification of economic activity into expansions and recessions refers to its cyclical component – broadly speaking, the behavior of the economy around its secular trend. In a stable economy like the U.S., it does not seem controversial to examine the growth rates (month-on-month or year-on-year) of the set of coincident economic indicators used by the BCDC. This method implicitly assumes a constant growth path and does not require specific modelling of the trend process. However in macroeconomics it is common to investigate business cycle phenomena by applying some filtering method to the raw data in the levels. We find this problematic for several reasons: (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 are likely to differ; and (4) common filtering methods often introduce additional and unwanted dynamic elements into the cyclical component.

However and for the sake of completeness, in this section we examine the classification skill of the BCDC for monthly and yearly annualized output growth rates, as well as deviations from a Hodrick and Prescott (1997) trend (HP); a Baxter and King (1999) trend (BK) where the cycle is defined over frequencies between 6 to 32 quarters; and from estimates of potential output reported by the Congressional Budget Office (CBO). The sample begins March 1947/February 1948 (depending on the monthly/yearly growth transformation) and ends November 2007, the month prior to the last release available from the BCDC. Arguably, we could have extended the sample further by assuming that for several periods after December 2007 the economy was in recession, but we prefer to abstain from speculation. Figure 4 displays the growth rates of these trends to get a sense of their variation over the business cycle.

Two basic results stand out. First, there is very strong conformity across the HP, BK and CBO trends, with some slight differences between CBO on one side, and HP and BK on the other. For example, in the 2001 recession, HP and BK trend output are declining whereas CBO potential output is increasing. Second, although some of the time trend-output grows during recessions, recessions generally coincide with periods in which trend output is low. This observation affects the timing of turning points as we show below.

Table 1 reports $AUROC$ estimates of cyclical GDP using these five detrending methods and the definition of recession provided by the BCDC, that is, a recession is the period from peak to trough, both months included (the next section explores the definition of what a recession is in more detail). It is immediately clear from Table 1 that while year-on-year GDP growth has the highest $AUROC$ (at 0.98 and virtually indistinguishable from the ideal value of 1), this maximum is achieved by shifting the beginning and end of recessions by three-months with respect to the BCDC chronology. Instead, month-on-month GDP growth matches the BCDC chronology, but because it is a noisier measure its $AUROC$ is only 0.89, high but statistically inferior (even at the 1% level) to the year-on-year GDP growth. Cyclical measures obtained from HP, BK and CBO trends have similarly inferior $AUROC$ values but their maximum is attained with a six-month phase shift with respect to the BCDC chronology. We think the explanation can be found in Figure 4, which shows that as the economy recovers from a recession, both trend and cycle improve but the distance between trend and cycle remains sufficiently large to delay the timing by which an exit from the recession is detected.

### 3.2 Four Definitions of 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,[10] 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

[10] www.nber.org/cycle/

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. In addition and because there could be phase shifts across indicators (for example, it is well-known that employment tends to lag considerably in economic recoveries), we examine a window $h$ of $\pm 24$ months around turning points. At each horizon $h$, we calculate the corresponding $AUROC$ and denote the horizon at which the maximum $AUROC$ occurs as $h^*$. We do this for month-on-month and year-on-year growth rate transformed data for completeness.

The BCDC claims to base its decisions on 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). Consequently, we constructed these indicators with data obtained directly from the sources listed by the BCDC (more details are provided in the appendix). To allow for direct comparison between the quarterly and monthly indicators, we construct monthly interpolated series of GDP and GDI using the linear interpolation method described by the BCDC. We then take annualized month-on-month and year-on-year log differences to compute growth rates that we express in percentages to facilitate easier comprehension.

Table 3 reports $AUROC$ estimates for each indicator using the two growth transformations (month-on-month and year-on-year) for dates lining up with the BCDC chronology ($h = 0$) and for the phase shift that would maximize the value of the $AUROC$ ($h^*$). Generally speaking, either BDCD-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 BCDC chronology exactly or with a one month difference in all the cases. 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. 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.

### 3.3 The BCDC versus Statistical Dating Rules

The BCDC's dating of business cycles is held as the universally accepted gold standard 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 an interpolated version of Chauvet and Hamilton's (2005) quarterly Markov-switching smoothed transition state probability index (henceforth the CH index), which is readily available[11] and transparent. The two-state Markov chain specified in the model captures the underlying unobserved state of whether the economy is in expansion or recession. In order to translate the transition state probabilities into a monthly zero-one indicator about the state of the economy, we interpolate the quarterly index linearly and then apply the simple rule-of-thumb that any period with a recession probability greater than a given threshold value is classified as a recession. In order to find the optimal threshold, we performed a grid-search over the space 0.5-0.9 and found 0.75 to maximize the $AUROC$ for the majority of indicators analyzed here.

Chauvet and Piger (2008) produce a similar index,[12] and which we will denote CP. CP first estimate a dynamic factor model in the vein of Stock and Watson (1989) using data on four coincident variables: nonfarm payroll employment, industrial production, real manufacturing and trade sales, and real personal income less transfer payments. The common factor $\mu$ is assumed to follow the process $\mu = \mu_0 + \mu_1 S_t$, where $S_t$ is an unobserved latent

[11] `www.econbrowser.com/archives/rec_ind/description.html`
[12] Available from Jeremy Piger's homepage at `www.uoregon.edu/~jpiger/`

variable about the state of the economy. Estimation of the model produces an estimate of the probability that the economy is in recession. CP use a two-step process to then translate this probability into a binomial variable. First, CP record when does the estimated probability become greater than or equal to 0.80 for three consecutive months. These dates are classified as recession. Let the first month of this series be month $t$. Then the beginning of the recession is dated as the first month prior to month $t$ for which the probability of recession is greater than 0.5.

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 with month-on-month and year-on-year growth transformations. The bottom panel repeats the exercise but allows for a phase shift $h$ in the range of $\pm 24$ months. We then report the maximum $AUROC$ achieved and the corresponding $h^*$.

Broadly speaking, we find that BCDC-PI generates the highest overall $AUROC$ values regardless of how one looks at the data although the differences tend to be relatively minor in most cases (in the practical as well as in the statistical sense). The results are interesting, specially since the CP index combines information from several of the coincident indicators used by the BCDC and optimally allocates the data into the two distributions in the mixture. However, the BCDC's dating process appears to classify a broad range of variables while sacrificing very little by means of misclassification for any individual series.

# 4 Indices of Business Conditions

The analysis in Section 3 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 released 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 along similar lines.

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.[13] The second index included is 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.[14]

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 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[15]). We also include the index that we introduced in Section 2 based on a standardized measure of the counts of 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.

We evaluate these indices with the most recently available data vintage since real-time vintages are not available for a long enough period. 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.

The results of this analysis 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

[13] www.chicagofed.org/economic_research_and_data/cfnai.cfm
[14] www.philadelphiafed.org/research-and-data/real-time-center/business-conditions-index/
[15] www.ism.ws/ISMreport/content.cfm?ItemNumber=10752&navItemNumber=12961

$AUROC$ 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 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:

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

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.

# 5 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 still maintain the working convention that the BCDC's chronology is the "gold standard" that these predictions should try to 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.

## 5.1 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 $[1, 0.5]$. 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.

### 5.2 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.

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.

## 6 Discussion

Cyclical fluctuations of economic activity have long been categorized into expansions and recessions in implicit recognition that the economy evolves differently in each state. Policy-makers may not be as concerned with momentary lapses into economic weakness as they may be with full transitions into the recessionary state even when economic data remains relatively benign. This paper offers fresh views on the problem of classifying economic activity into expansions and recessions. In fact, the methods that we explore are well suited to make explicit the trade-offs that a policy-maker faces in characterizing the business cycle.

To our knowledge, we are the first to provide a direct measure of the quality of the chronology of business cycles provided by the BCDC. This is important because we are able to assure researchers and the public that the chronology has considerable classification value, even when compared to statistical models tailored to optimize how the data should be categorized. Furthermore, our analysis yields insight into the timing of these transitions: year-on-year growth rates of economic variables can be classified more accurately but this requires shifting the beginning and end of recessions by three-to-four months. Employment cycles would have to be shifted by an additional three-to-four months.

In order to design an effective policy response one must determine what is the current state of the economy and when are future transitions expected to occur. Business conditions indices maintained by the Federal Reserve Bank's of Chicago and Philadelphia provide accurate signals in real time. The components of the Index of Leading Indicators provide fairly accurate prediction-classification up to horizons of one year (with an $AUROC$ close to 0.9 throughout) but even two-years out one can do better than a coin-toss. Here a novel observation is that no single linear combination of the components of the ILI is likely to work well since we have uncovered strong variation across horizons and in the manner in which each of the components help classify future turning points.

We conclude by noting that understanding the difference between classification ability and model fit is important. In the usual least squares scenario, model fit improves when the Euclidean distance between an observation and the regression line is made small, regardless of the sign of the regression error. Therefore, extreme observations tend to drive the slope of the regression line that is estimated. However, in a classification scenario the sign of the regression error is much more important – extreme events are easily assigned to the correct class but it is much more difficult to assign observations in the neighborhood of the regression line. Tilting of the regression line due to extreme observations can therefore result simultaneously in better fit but worse classification. For this reason, there are a number of new statistical methods tailored for classification, such as linear or quadratic discriminant analysis, neural networks, support vector machines and boosting algorithms; techniques that will surely permeate into economics applications. We hope to have provided a new road-map to explore the taxonomy of business cycle phenomena with these more sophisticated techniques and to have provided methods that can be easily implemented and interpreted even by a non-technical audience.

# 7 Appendix

## 7.1 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.

# 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.

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.

Jordà, Òscar and Alan M. Taylor (2009a) "The Carry Trade and Fundamentals: Nothing to Fear but FEER Itself," U.C. Davis, NBER working paper 15518.

Jordà, Òscar and Alan M. Taylor (2009b) "Investment Performance of Directional Trading Strategies," U.C. Davis, *mimeo.*

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

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.

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.

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.