Towson University
Department of Economics
Working Paper Series

Working Paper No. 2016-15

Predicting U.S. Business Cycle Turning Points Using Real-Time Diffusion Indexes Based on a Large Data Set

by Herman Stekler and Yongchen Zhao

September 2016

Predicting U.S. Business Cycle Turning Points Using Real-Time Diffusion Indexes Based on a Large Data Set

Herman Stekler[a] and Yongchen Zhao[b 1]

[a] *Department of Economics, George Washington University*

[b] *Department of Economics, Towson University*

Abstract

This paper considers the issue of predicting cyclical turning points using real-time diffusion indexes constructed using a large data set from March 2005 to September 2014. We construct diffusion indexes at the monthly frequency, compare several smoothing and signal extraction methods, and evaluate predictions based on the indexes. Our finding suggest that diffusion indexes are still effective tools in predicting turning points. Using diffusion indexes, along with good judgement, one would have successfully predicted or at least identified the 2008 recession in real time.

Keywords: forecasting recession; real-time data; probability forecast

JEL Codes: C43, C53, C55, E37

---

[1] Herman Stekler (hstekler@gwu.edu); Yongchen Zhao (yzhao@towson.edu). The authors thank David Small of the Federal Reserve Board for kindly providing the data set used in this study.

Predicting U.S. Business Cycle Turning Points Using Real-Time Diffusion Indexes Based on a Large Data Set

1. Introduction

This paper revisits an old topic - the ability to forecast cyclical turning points using the diffusion index, in this case, of a real-time large data set. Diffusion indexes have long been used in both business cycle analysis and in macroeconomic forecasting. A diffusion index is simply a number - the percentage of series from a given sample that are increasing over a specified time period. In an ex post cyclical context, Moore (1950) showed that the *historical* diffusion indexes led the turns in the business cycle peaks by about eight months but this finding follows by definition.

This index is derived from historical data and is based on the number of series which have not yet reached their cyclical peaks (troughs) and not all series reach their peaks (toughs) simultaneously.[2] Because the NBER cyclical peak (trough) is defined as the date by which the largest number of the series had their specific turning points, the turns in the *historical* diffusion index must lead the economy's turns. In real-time, however, a forecaster does not know whether a particular series has reached a turning point. For this approach to be useful in forecasting economic activity it had to be based on information available in real-time.

[2] Historical diffusion indexes come from the series which are called rising or falling only with reference to each series' specific cyclical turns. Thus a series is considered to be rising during every month between its trough and its subsequent peak.

Alexander (1958) was the first to test the predictive powers of a *pseudo-real-time* diffusion index, which, in his article, was derived from the components of the Federal Reserve Board (FRB) Index of Industrial Production.[3] The first step was to calculate the sign of the month-to-month differences of each of the components of the FRB Index regardless of the stage of the cycle. This determined whether that component was increasing based on the data available at that time. The percent of the series that were expanding in real time was then calculated, yielding a real-time diffusion index. The rationale for this approach is that the number of series that are expanding should decline as the economy slows and then enters a recession. Thus it will have a peak prior to the beginning of a recession.

Because the index was very volatile, Alexander next smoothed this index.[4] The final step in using a real-time diffusion index for forecasting is to select an explicit rule for determining when a signal of a turn has occurred. While he did not develop this explicit rule, Alexander concluded that a pseudo-real-time version of the diffusion index predicted all of the cyclical peaks between 1919 and 1956, but that it also generated a significant number of false predictions.

Ad hoc rules for identifying the signals were developed by Alexander and Stekler (1959) and enabled analysts to determine (1) whether turning points were predicted, (2) the forecasting lead, and (3) the number of false signals. The original rule was based on the number of months that the indicator was below (above) a peak (trough). If every decline from a peak is counted as a

---

[3] It is pseudo real time because historical data are used, but each data point is treated as if it were observed for the first time.

[4] Alexander's smoothing formula corresponded to fitting a second degree parabola.

signal, it was shown that there would be a very large number of false predictions. Consequently, the use of an "*n* or more months up or down" rule was suggested. This rule is an implicit smoothing device that generates the tradeoff between the number of false turns and the average forecasting lead.[5]

More systematic approaches for evaluating indicators have been developed but have never been used to evaluate diffusion indexes. For example, probits are used to calculate the probability that an indicator has signaled a turning point and the indicator is then evaluated using quadratic probability scoring rules, some of its attributes, or Receiver Operating Characteristics (ROC) curves. In fact, since the 1960s there have been few attempts to forecast turning points using real-time diffusion indexes.[6] This is a surprise since the failure to forecast turning points is one of the major concerns of macroeconomic forecasting and we know that diffusion indexes can forecast those turns, albeit with false signals. In fact, one of the components of the Conference Board's Composite Index of Leading Indicators is a diffusion index.[7]

---

[5] Vaccara and Zarnowitz (1977) suggested an alternative rule for identifying a predictive signal from an indicator. There must be three consecutive declines in the indicator before saying that a signal has occurred. There is a distinction between the "three month down" and the three month consecutive decline rules. The indicator may be below peak for three months but may not have declined in every one of those three months.

[6] Stock and Watson (2002) entitled their paper dealing with transfer functions as "Macroeconomic Forecasting Using Diffusion Indexes". However, while the transfer functions use the data from all the series, they are not diffusion indexes as conventionally defined.

[7] It is the Institute of Supply Management's New Order Index, which, "reflects the number of participants reporting increased orders during the previous month compared to the number reporting decreased orders, and this series tends

This paper will therefore construct a diffusion index from a database that consists of more than 150 series for the period January 1982 to September 2014. We will then determine how well the index forecast the cyclical turns that occurred in this period. The data and the methodology that we use are explained in the next section. The third section discusses nowcasting and forecasting and evaluates the forecasts. Concluding remarks are offered in the last section.

2. The Real-Time Data Set and Index Construction

2.1. The large real-time data set

The diffusion indexes are constructed from the series in the Giannone *et al.* (2008) data set. That data set contains 183 monthly economic series for the period January 1982 – February 2005 and refers to more sectors than were the focus of previous studies. The earlier studies primarily focused on industrial production indexes or closely related real income/output measures such as personal income or sales[8]. This data set also includes indicators on price levels, interest rates, exchange rates, employment situations, bank and money reserves, and indicators from business

---

to lead the business cycle". Moreover the Federal Reserve publishes diffusion indexes constructed from the components of its Index of Production. For more information on the Conference Board's index, see https://www.conference-board.org/data/bci/index.cfm?id=2160.

[8] See Stock and Watson (2014) and references therein.

outlook surveys.[9] We only use 166 of these series because we exclude the exchange rate and money stock variables in this study.

The data set can thus be visualized as formed from two distinct parts. From January 1982 to February 2005, all of the data are the revised historical numbers that were available in March 2005 and are treated as pseudo-real-time observations. Beginning in March 2005, the data set is updated every Friday. Each time the data are updated, a new data file is created that contains the entire history of all the indicators updated to reflect both the latest information and any revisions to previous observations that may have occurred.

It is possible to compare the difference of each variable between data files. If there were no new information or revisions of a particular variable, the values for that variable would be the same in the two files. On the other hand, there may be a difference between the two files and this may occur for two reasons. If a new release of the indicator becomes available between the two Fridays, there will be one more observation of the indicator in the more recent data file. If revisions to previously released values are announced between the two Fridays, the more recent data file will contain the revised values while the earlier data file contains values before the revisions. By comparing successive data files, one recreates the stream of real-time data flow. As we discuss in the following subsections, this data stream is the basis for our index construction.

[9] For a detailed list of series contained in the data set, as well as their publication schedule, please see Giannone *et al.* (2008). The appendix of the paper also provides a detailed list of data transformation methods used to treat each indicator.

As noted, both data release schedules and data revisions create differences between data files. The former factor also result in the data files having "jagged edges", as publication lags vary among the indicators. This effect is taken account in the post March 2005 real-time data and poses no problem for our analysis. However, since there is only one recession after March 2005, evaluating the performance of the indexes solely based on data from the post-2005 period is insufficient. Using the data beginning in 1982, we can obtain a much longer time series, merely by creating a series of pseudo-real-time data sets which are obtained by recreating the jagged edges derived from the latest data file.[10] This assumes that the publication dates of the various variables have not changed between 1982 and 2014. Therefore, both the pseudo-real-time data and the true real-time data are used to construct the diffusion indexes for the entire period, 1982-2014.[11]

Two adjustments must be made before it is possible to construct a diffusion index. First the data are expressed in levels and must be transformed to induce stationarity. We use the Giannone *et al.* (2008) procedure.[12] Depending on an indicator's time series properties, if a transformation is needed, the indicator may be transformed to a monthly growth rate, or a monthly difference, or a monthly difference of the annual growth rate.

Second, we must ensure that the indicators conform to a common cyclical pattern, i.e. they increase when the economy is growing, etc. It is thus necessary to determine whether the series

---

[10] This approach has been adopted by Lahiri *et al.* (2015).

[11] When we present the results below for the entire period, we refer to this series as the pseudo-real-time series, it is based on vintage 2014 data.

[12] Since lagged values are used when transforming the data, the data from the first year, 1982, that remain after the transformation are discarded, so that all the indexes constructed from the transformed data begin in 1983.

are pro-cyclical or counter-cyclical. Once we have determined this cyclicality, we assign a different value to the signal obtained from a pro-cyclical indicator to that yielded by a counter-cyclical indicator. This is done by simply multiplying the values of the counter-cyclical indicators by -1. In the data set, the six unemployment and housing supply variables are considered counter-cyclical. All other variables are treated as pro-cyclical.[13]

2.2. Index construction

In order to construct the diffusion index, it is necessary to determine the number of series that are increasing (decreasing). Since all the indicators in the data set have been differenced or transformed to growth rates, an indicator increased (decreased) from the previous month if the sign of the change, $x_t - x_{t-1}$,was positive (negative). When the value is zero, the indicator is counted as unchanged.[14] The diffusion index, for a given month, is then defined as the percentage of series increasing minus the percentage of indicators decreasing plus 100. If the economy were expanding (contracting), more indicators would presumably be increasing (decreasing) and the index would be greater (less) than 100. Once we obtain the monthly index, a quarterly index can be constructed as an average of the monthly values.[15]

---

[13] In addition to relying on intuition, we examined regressions of real GDP growth rate and the binary variable indicating NBER recession on each variable in our data set.

[14] In our data set, on average, only three indicators stay unchanged from one month to the next.

[15] Alternatively, one could simply use the last monthly value of a quarter as the quarterly index, or generate the quarterly index by comparing data from the last week of a quarter to that of the previous quarter. We argue that both alternatives result in information loss, as data from the first and second month of a quarter go unused. We focus on

A number of important characteristics of this approach should be noted. First, on any given Friday, the most up-to-date information is used to calculate the difference or growth rate of each indicator. Suppose at time period $t + 1$, the value x_t is newly released, and the value x_{t-1} is revised to be x^*_{t-1}. Our procedure calculates the difference as $x_t - x^*_{t-1}$.[16]

Second, the data set is based on all the information that is available at a given time for each time series. Given the differences in publication lags, observations for different months are pooled together. Suppose the index for May 2007 is 110. This does not mean that the percentage of indicators that *actually* increased in May was 10% more than that *actually* decreased. Instead, it means that out of all the new data releases/revisions *announced* in May, the percentage of increases is 10% more. Some of the actual announcements are about May, some about April, and some about even earlier months, depending on the mix of publication lags.

Third, in constructing the index, all of the indicators are weighted equally. This is the procedure that has been used in the past, but it can be argued that different indicators should be weighted differently in the index, depending on the composition of the data set.[17] The weighting

---

the monthly index in subsequent discussions. Since the quarterly index is by definition a much smoother version of the monthly index, it gives fewer false signals but (obviously) at a lower frequency.

[16] This is in general different from calculating difference or growth rate using the first vintage (initial release) of an indicator when the indicator is subject to revisions, i.e. $x_t - x_{t-1}$.

[17] Stock and Watson (2014) considered both unequal weights and lag adjustments. Since we are not particularly interested in the exact date of turning point, lag adjustments are not necessary. Also, given the nature of the stream of real time data, lag adjustments necessarily results in loss of information given data publication lag, defeating the purpose of a real time index. Giannone *et al.* (2008) sorted the variables in the real time data set into several blocks.

issue is similar to the one in the forecast combination literature which indicates that a weighted average constructed using estimated weights often produces forecasts that are inferior to a simple average.[18] We, therefore, continue to assign equal weight to all our indicators.

2.3. Smoothing

The previous literature has shown that diffusion indexes as constructed are very volatile, and it is, therefore, necessary to smooth the index. Our results confirm these findings. However, smoothing introduces a lag, which partially offsets the benefit of a real-time index. So the smoothing method must be carefully selected to provide a balance between smoothness and lag.

Since the indexes are constructed with the purpose of nowcasting or forecasting in mind, certain features should be considered in selecting a smoothing approach. First, the method should be easily implemented in real time. Second, smoothing should introduce a minimum amount of lag or "phase shift". In addition, a desirable method should be computationally efficient and

---

We experimented with creating block-level diffusion indexes and weighting different blocks differently when aggregating them into a single index. Our in-sample analysis did not show significant improvement of this weighted index over the unweighted index. Therefore, we do not consider weighting and lag adjustment in this paper.

[18] Zhao (2016) shows that estimating even a set of five weights using a short time series in real-time may produce combined forecasts that are inferior to simply averaging the individual forecasts. This is especially true when the target variable or individual's forecasts are subject to structural breaks.

intuitively straightforward. Ideally, as new data are obtained, the next smoothed value should be estimated without needing to revise all previous estimates.[19]

In an early study, Alexander (1958) smoothed the diffusion indexes by fitting a second-degree polynomial to the data by using a rolling window of seven observations. The second to last observation was then used as the value of the smoothed index. Limited by the capability of computing devices that were then available, this approach provided reasonable performance without introducing too long a lag. Given the computing power now available we consider other approaches as well. In a recent survey, Alexandrov *et al.* (2012) review several modern approaches to the problem of trend extraction, including a model based approach, nonparametric linear filtering, singular spectrum analysis, and wavelets. However, as will become evident in subsequent sections, it is not the level of sophistication of the smoothing method that determines the forecast accuracy of the index. Simple smoothing devices, such as moving average, may deliver good results when coupled with an appropriate signal extraction method and/or an ad hoc rule.

In this study, a few common smoothing methods are compared. The starting point is the polynomial fitting used in Alexander (1958). Let the unsmoothed value of a diffusion index be y_t and the smoothed index be s_t. Alexander's method uses the smoothing formula $14s_t = 5y_t + 4y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4} - y_{t-6}$. The second method is simple exponential smoothing. After comparing several alternatives, we set the smoothing parameter to be 0.5. The third

[19] If one desires revising the smoothed index values each period, the before-smoothing index values, not the smoothing method, are what ought to be updated, since revisions to the underlying indicators would change the index numbers before any smoothing is implemented.

smoothing method is a moving average. When smoothing the quarterly indexes, a 4-quarter moving average is used, whereas a 6-month moving average is used when smoothing the monthly indexes. We also consider the Hodrick–Prescott (HP) filter, where the naïve choice of $\lambda = 14400$ is used for our monthly indexes and $\lambda = 1600$ is used with quarterly indexes. All the smoothing methods are implemented in real time using an expanding window with an initial size of 12 months, where applicable.[20] While we consider several methods, we do not intend to reach a specific conclusion as to which method is "correct" or "best". Rather, we recognize that the choice of a smoothing method is an empirical one and the best choice may well be time-varying. Our objective is to demonstrate the usefulness of the index itself and to show examples of the tradeoffs between the smoothness and the lag length.

2.4. Signal extraction

Consider a forecasting exercise where the target variable is the date of the turning point of a recession, as defined by the NBER. A forecaster collects data in real time and produces/revises a set of forecasts of this target variable on the last Friday of each month when the data set is updated. Given the smoothed values of the diffusion index, it is then necessary to determine when this index produces a signal indicating the onset of a recession. We consider two approaches to signal extraction. As discussed below, the first is a rule-of-thumb-type approach derived from the

[20] Several additional smoothing methods were examined and evaluated, including additional polynomial smoothing methods, alternative exponential smoothing methods and filters, such as the asymmetric band pass filter of Christiano and Fitzgerald (2003). The results are either similar to or worse than those obtained using the four methods reported here. Therefore, we omit these additional results but they are available upon request.

literature. The second is a regression-based approach. Both approaches can be used in identifying recession months/quarters and predicting business cycle turning points.

2.4.1 Simple decision rules

The first approach, based on simple decision rules, is easy to implement in real time and has been popular in the literature. A signal is identified whenever the index is below some pre-determined threshold. In practice, the threshold can be set to any value depending upon the desired tradeoff between the rate of successfully predicting a turning point (i.e., the hit rate) and the rate of predicting a turning point when none occurred (i.e., the false alarm rate). The optimum threshold should also depend on the chosen smoothing method.

Since the indexes are constructed as balance statistics, the theoretical neutral point is 100, where the percentage of increasing indicators equals that of decreasing indicators. If one used this threshold, the turn would only be identified after it occurred because, by definition, a recession would have already started once the index falls below 100. Therefore, in practice, the threshold that will be used will deviate from this theoretical value. An increase (decrease) in the threshold in order to increase (decrease) the lead time in the forecasts will induce a higher (lower) false alarm rate. We illustrate this issue in the next section by considering three thresholds: 100, 110, and 120.

In addition to these “below a threshold”-type rules, there are the ad hoc rules for identifying turning points that were discussed above. Specifically, we consider the decision rule suggested by Vaccara and Zarnowitz (1977). In this case, a signal of a peak is identified whenever the index shows three consecutive declines.

2.4.2 Probit models

Alternatively, using the index in a probit model, probability forecasts are generated. Using the same notation as in the previous section, the model can be written as

$$\Pr(I_{t+h} = 1) = \Phi\left(\alpha + \sum_{j=0}^{k} \beta_j s_{t-j}\right)$$

where I_t is the binary indicator of month t being a recession month and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Two specifications are considered here. In the first specification, we regress the binary indicator of recession on the smoothed index, and $k = 0$. The right-hand-side variable is lagged by h periods, where h is the forecast horizon, starting from $h = 0$ in the case of nowcasts. The second specification is otherwise identical to the first, except that we add two additional lags of the index to the right-hand-side, so that $k = 2$. The estimated models are then used to generate probability forecasts, which can be evaluated using metrics such as the Brier's (1950) quadratic probability score (QPS)[21]. Once the probability forecasts are available, it is easy to obtain binary signals using thresholds that provide the tradeoff between the hit rate and the false alarm rate. Since the optimum tradeoff obviously depends on specific applications, we do not attempt to suggest any specific threshold here. Instead, we plot the forecasts and report the area under the ROC curve[22], which is commonly referred to as the AUC statistic.

[21] The QPS is the probability forecast analogue of the well-known mean squared error (MSE).

[22] See Hanley and McNeil (1982) for a discussion of the ROC curve and the use of the AUC statistic.

2.4.3 Limitations

There is a practical limitation to the analysis of the probit forecasts due to the relatively short time series at our disposal. Our real-time data set contains only the 2007-2009 recession. Two additional recessions (1991 and 2001) happened during the longer sample period for which the pseudo-real-time data are available. We can produce out-of-sample forecasts for the 2005 to 2014 period by using 1984 to 2004 data as the estimation sample. The same set of estimated parameter values are used throughout 2005 to 2014 without updating. In this scenario a forecaster would be trying to predict the (now known) Great Recession based on information from the two previous recessions.[23] However, this limitation does not exist when the rule-of-thumb signal extraction method is used, as no estimation or prior data is needed in the process. Therefore, binary forecasts made in this fashion can be evaluated using the entire sample from 1984 to 2014.

3. Evaluation of the Monthly Diffusion Index[24]

This section evaluates the performance of the forecasts obtained from the monthly diffusion indexes in two stages. We first examine the forecasts from the pseudo-real-time index based on the 2014 vintage data. This is the procedure that an individual who only had that data would use in determining whether, with hindsight, this diffusion index generated accurate forecasts of

[23] It is worth noting that this empirical strategy puts a more stringent test on our indexes compared to a setup where a series of one-step-ahead out-of-sample forecasts are used. Because in our exercises, forecasts are based on a much smaller information set – one that only extends to 2004.

[24] We only present the results for the monthly index because the results for the quarterly index are virtually identical with the exception that the quarterly index is much smoother. These data may be obtained from the authors.

cyclical peaks. However, we also have real-time data beginning in 2005 that enable us to evaluate the index's performance in predicting the beginning of the Great Recession and assess how well it worked in practice.

In each stage we first discuss the techniques that were used to smooth the data and then discuss the performance of the smoothed indexes in discriminating between recessionary and non-recessionary months and in forecasting turning points. Stekler and Ye (2016) emphasize that there is a difference between (1) classifying a month as recession/non-recession and (2) predicting the onset of a recession. Our evaluation considers both of them. For the purpose of classification, we resort to the standard statistics – the QPS and the area under the ROC curve (AUC). For the purpose of predicting the onset of a recession, we directly report the number of correct predictions and false positives and also present graphical diagnostics.

3.1. Pseudo-real-time-index

3.1.1 Comparison of smoothing procedures

Figure 1 shows the indexes before smoothing (dots) and presents the results of alternative smoothing methods (solid lines). We observe that all four smoothing methods introduce a similar amount of lag. The HP filter produces indexes with the highest level of smoothness and may seem to be the most effective method. On the other hand, as discussed below, classification accuracy depends heavily on the signal extraction method as well. So smoothing procedures should not be evaluated in isolation. Rather, one should judge the forecasting procedure as a whole.

3.1.2 Classification accuracy

One of the criteria for judging the performance of our diffusion index is how well it discriminates between recessionary and non-recessionary months. The results for the entire period, 1984-2014, obtained by combining the various simple decision rules in conjunction with the alternative smoothing methods are presented in Table 1[25]. The first finding is that the Brier Scores (QPS) are a function of the smoothing method that was employed. As expected they are very dependent upon the magnitude of the threshold. The "100-threshold" clearly is the best in distinguishing between these two categories. When an equal number of series are expanding and contracting it is likely that the economy is approaching a cyclical turn. This was the reasoning that motivated Alexander (1958) to propose this method as a technique for forecasting turning points. Our empirics indicate that this approach is still valid in a nowcasting environment. The second point to note is that with one exception, the "100-threshold" rule has a lower QPS than the "three month consecutive decline" rule.

Although the probit models are estimated using data up to 2004, their forecasting performance is based on pseudo-real-time data only from 2006-2014. The results obtained from the probits that were derived from pseudo-real-time data are presented in Table 2. The findings indicate that the QPS and AUC statistics again depend upon the method of smoothing that is employed. In this case, the ability to classify the months into the two categories declines with the length of the forecasting lead. The errors are relatively low and the AUC is high even at the longer horizons. As a benchmark, a naïve constant forecast of 0.15, which corresponds to the proportion of recession periods in the data, has a QPS of 0.13. The forecasts up to 3-month perform better than the naïve benchmark. These forecasts also have much higher AUC values than 0.5 (the AUC

[25] As mentioned before, data transformation and the HP filter together each uses 12 months of data. Our data set starts from 1982, but our evaluation sample starts from 1984. This is also the reason that real time forecast evaluation is conducted starting March 2006, rather than March 2005.

value associated with no-skill forecasts). These results are further evidence that the diffusion index is able to classify the months into the two categories.

3.1.3 Forecast accuracy: cyclical turns

Using alternative signal extraction rules, we next ask whether the diffusion index predicts all three recessions without generating an excessive number of false signals. Figure 2 compares the results obtained from the alternative signal extraction methods. In each plot, from top to bottom, we show the predictions made using the "below a threshold" rule with thresholds of 100, 110, and 120, and the "three consecutive declines" rule. A signal is shown as a bar with dark shade. The actual recession months are lightly shaded.

Table 3 presents the forecasting lead and the number of false signals that are associated with each of the signal extraction rules. Two signals with no more than six months in between are counted as one.[26] Note that the count of false signals in Table 3 does not account for the "length" or "strength" of the signals. For example, a false signal that lasts 12 months is much stronger than one that lasts only one month. To address this issue, we report the average length of false signals in the last row of the table. As expected, when using the "below a threshold"-type decision rule, there is a tradeoff between the size of the forecasting lead and the number of false positives. Again the results depend on the smoothing formula that is utilized, but the rules predict all three recessions with differing lags and numbers of false signals. As expected, a higher threshold provides a longer forecast lead but creates more and stronger false signals.

---

[26] The desired choice here depends on the particular application. The results in Table 3 are based on what Figure 2 shows. So from Figure 2 one could easily see what the results are likely to be if an alternative choice is preferable.

In this time period, the “110 threshold” rule gives the best results. The signals arrive with an average of 6 to 10 months lead, and there are only 3 or 4 false signals.[27] We should, however, note that the threshold rules continue to predict a recessionary period for many months after the recession ended, but the three months consecutive declines” rule does not yield this specific result.

As for the probits, the first graph in Figure 3 shows that the probability that a recession would occur in 2007-2008 was extremely low even as the US had already entered the Great Recession. The probability of a recession did not exceed 50% until the recession was well underway. These results indicate that despite being able to classify the months, the probits did not yield good forecasts.

3.2. Real-time index

We now seek to determine whether the diffusion index would have yielded the same results in a real-time environment as were obtained from the use of the pseudo-real-time numbers. In this analysis we use only the real-time data that were available between 2006 and 2014.

3.2.1 Smoothing the index

We apply the same smoothing formulas to the real-time index that were used in conjunction with the pseudo-real-time data. Figure 1 again shows that the HP filter yields the smoothest series, at the cost of a somewhat longer lag.

---

[27] We should note that even if a signal arrives with a small lag that would be useful information in identifying a recession because the NBER does not indicate that a recession has begun for many months after the event.

3.2.2 Classification accuracy

When we use real-time data, we obtain results that are similar to those that were generated by the pseudo-real-time numbers. The results still depend on the smoothing method that is used; the QPS increases when the threshold is increased; the below “100 threshold” has a smaller QPS than the “three consecutive months” rule.

The main difference in the results is that the QPS of the real-time data is less than that of the pseudo-real-time figures for the entire period. However, the relevant question is how the real-time and pseudo-real-time results compare for the same time period, 2005-2014. As Table 1 shows, the QPS of the pseudo-real-time data is generally lower, especially when the below “100 threshold” or the “three consecutive months” rule is used. This is expected, since data revisions should improve data accuracy and therefore classification accuracy. However, small revisions are unlikely to reverse the change of an indicator from its previous month’s value. Since most revisions are small, as we increase the threshold, the effect of data revisions diminishes. When the below “110 threshold” rule is used, there is little difference between the QPS of real-time data and pseudo-real-time data. When the “120 threshold” rule is used, the QPS of pseudo-real-time data is generally larger. The results for the probit regression are clear: the real-time data do not classify the months as well. The QPS is larger and the AUC is smaller for every smoothing procedure even though it is still superior to the no-skill naïve model. (Table 2).

3.2.3 Forecast accuracy: 2007-2009 recession

Using either the HP filter or a moving average formula and different thresholds, the diffusion index would have provided useful information because it either forecast the Great Recession or identified its existence within two or three months of its occurrence. This is a significant result because the NBER only established the date of the peak more than two years after it occurred. Moreover, Lundquist and Stekler (2012) and Stekler and Talwar (2013) showed that economists did not predict the recession in advance and experienced a considerable lag in even identifying it**.**

However, Figure 2 and Table 3 indicates that the use of real-time data in making binary predictions has somewhat degraded the performance of the diffusion index relative to the record obtained from the revised figures in the 2014 vintage data. Either the forecasting lead has decreased or the number of false positives has increased. Similar results can be observed in the probability forecasts. (Figure 3).

4. Concluding Remarks

In this paper, we consider whether it is possible to predict (or at least identify without long lags) U.S. recessions using real-time diffusion indexes based on a large data set. We construct the diffusion indexes, compare alternative smoothing and signal extraction methods, and evaluate nowcasts and forecasts made using the indexes. Our results show that the diffusion indexes, albeit their simple construction, remain effective tools for predicting business cycle peaks. More importantly, we find that the diffusion indexes would have performed well in forecasting the 2008 recession in real time. Our results also suggest that the most effective way of using the indexes is in conjunction with simple decision rules such as “below a certain threshold” and “three consecutive declines”.

References

Alexander, S. S. (1958): "Rate of Change Approaches to Forecasting--Diffusion Indexes and First Differences," *The Economic Journal*, 68, 288–301.

Alexander, S. S., and H. O. Stekler (1959): "Forecasting Industrial Production--Leading Series versus Autoregression," *Journal of Political Economy*, 67, 402–409.

Alexandrov, T., S. Bianconcini, E. B. Dagum, P. Maass, and T. S. McElroy (2012): "A Review of Some Modern Approaches to the Problem of Trend Extraction," *Econometric Reviews*, 31, 593–624.

Brier, G. W. (1950): "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, 75, 1–3.

Christiano, L. J., and T. J. Fitzgerald (2003): "The Band Pass Filter*," *International Economic Review*, 44, 435–465.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The Real-Time Information Content of Macroeconomic Data," *Journal of Monetary Economics*, 665–676.

Hanley, J. A., and B. J. McNeil (1982): "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, 143, 29–36.

Lahiri, K., G. Monokroussos, and Y. Zhao (2015): "Forecasting Consumption," *Journal of Applied Econometrics*, Forthcoming.

Lundquist, K., and H. O. Stekler (2012): "Interpreting the Performance of Business Economists During the Great Recession," *Business Economics*, 47, 148–154.

Moore, G. H. (1950): *Statistical Indicators of Cyclical Revivals and Recessions:* NBER.

Stekler, H. O., and R. M. Talwar (2013): "Forecasting the Downturn of the Great Recession," *Business Economics*, 48, 113–120.

Stekler, H. O., and T. Ye (2016): "Evaluating a Leading Indicator: An Application: The Term Spread," *Empirical Economics*, Forthcoming.

Stock, J. H., and M. W. Watson (2002): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business & Economic Statistics*, 20, 147–162.

——— (2014): “Estimating turning points using large data sets,” *Journal of Econometrics*, 178, Part 2, 368–381.

Vaccara, B. N., and V. Zarnowitz (1977): “How good are the leading indicators,” *ASAProBuEc*, 41–50.

Zhao, Y. (2016): “Robustness of Forecast Combination in Unstable Environment,” *Working paper*.

Table 1. QPS of Binary Forecasts: Pseudo-Real-Time, Jan 1984 to Feb 2006; Pseudo-Real-Time, Mar 2006 to Sep 2014; Real-Time, Mar 2006 to Sep 2014

Index and Smoothing Method	Below 100	Below 110	Below 120	Three consecutive declines
Pseudo Real Time (Jan 1984 to Feb 2006)				
Alexander (1958)	0.169	0.376	0.568	0.177
Exponential Smoothing	0.165	0.387	0.594	0.120
HP Filter	0.192	0.383	0.613	0.171
Moving Average	0.132	0.387	0.579	0.173
Pseudo Real Time (Mar 2006 to Sep 2014)				
Alexander (1958)	0.078	0.206	0.480	0.157
Exponential Smoothing	0.108	0.255	0.500	0.176
HP Filter	0.127	0.176	0.490	0.176
Moving Average	0.078	0.196	0.539	0.176
Real Time (Mar 2006 to Sep 2014)				
Alexander (1958)	0.108	0.235	0.451	0.245
Exponential Smoothing	0.127	0.255	0.441	0.206
HP Filter	0.137	0.196	0.412	0.162
Moving Average	0.088	0.167	0.520	0.196

Table 2. QPS and AUC of Probability Forecasts, Pseudo-Real and Real-Time, Mar 2006 to Sep 2014

Index and Smoothing Method	QPS					AUC				
	h=0	h=1	h=2	h=3	h=4	h=0	h=1	h=2	h=3	h=4
Pseudo Real Time										
Alexander (1958)	0.058	0.075	0.089	0.103	0.117	0.974	0.944	0.923	0.888	0.892
Exponential Smoothing	0.076	0.089	0.103	0.116	0.133	0.942	0.917	0.892	0.884	0.869
HP Filter	0.061	0.076	0.088	0.101	0.113	0.977	0.949	0.928	0.907	0.903
Moving Average	0.061	0.077	0.089	0.106	0.123	0.977	0.960	0.949	0.933	0.893
Real Time										
Alexander (1958)	0.078	0.094	0.103	0.117	0.129	0.940	0.920	0.900	0.849	0.865
Exponential Smoothing	0.094	0.103	0.115	0.128	0.142	0.912	0.891	0.846	0.854	0.830
HP Filter	0.084	0.099	0.109	0.121	0.132	0.928	0.903	0.887	0.856	0.850
Moving Average	0.086	0.099	0.106	0.121	0.137	0.955	0.937	0.929	0.898	0.851

Table 3. Forecasting Lead and Number of False Signals, Alternative Methods

Date of Peak	Lead (months)							
	HP Filter				Moving Average			
	Below 100	Below 110	Below 120	Three consecutive declines	Below 100	Below 110	Below 120	Three consecutive declines
Pseudo-Real-Time								
1990	-4	10	12	15	-4	12	13	0
2001	0	3	7	6	3	6	8	6
2007	-3	4	15	3	-2	13	15	3
Average Lead	-2.3	5.7	11.3	8	-1	10.3	12	3
Number of False Signals	1	3	3	8	4	4	4	12
Average Length of False Signals	9	11	31	3.6	2.3	11.5	23	12
Real-Time								
2007	-2	3	16	2	-2	3	15	2
Number of False Signals	0	3	1	1	0	2	2	3
Average Length of False Signals	N/A	2	12	1	N/A	7.5	13	3.3

Figure 1. Smoothed Monthly Indexes, Pseudo-Real-time, 1984 to 2014 and Real-Time 2005 to 2014

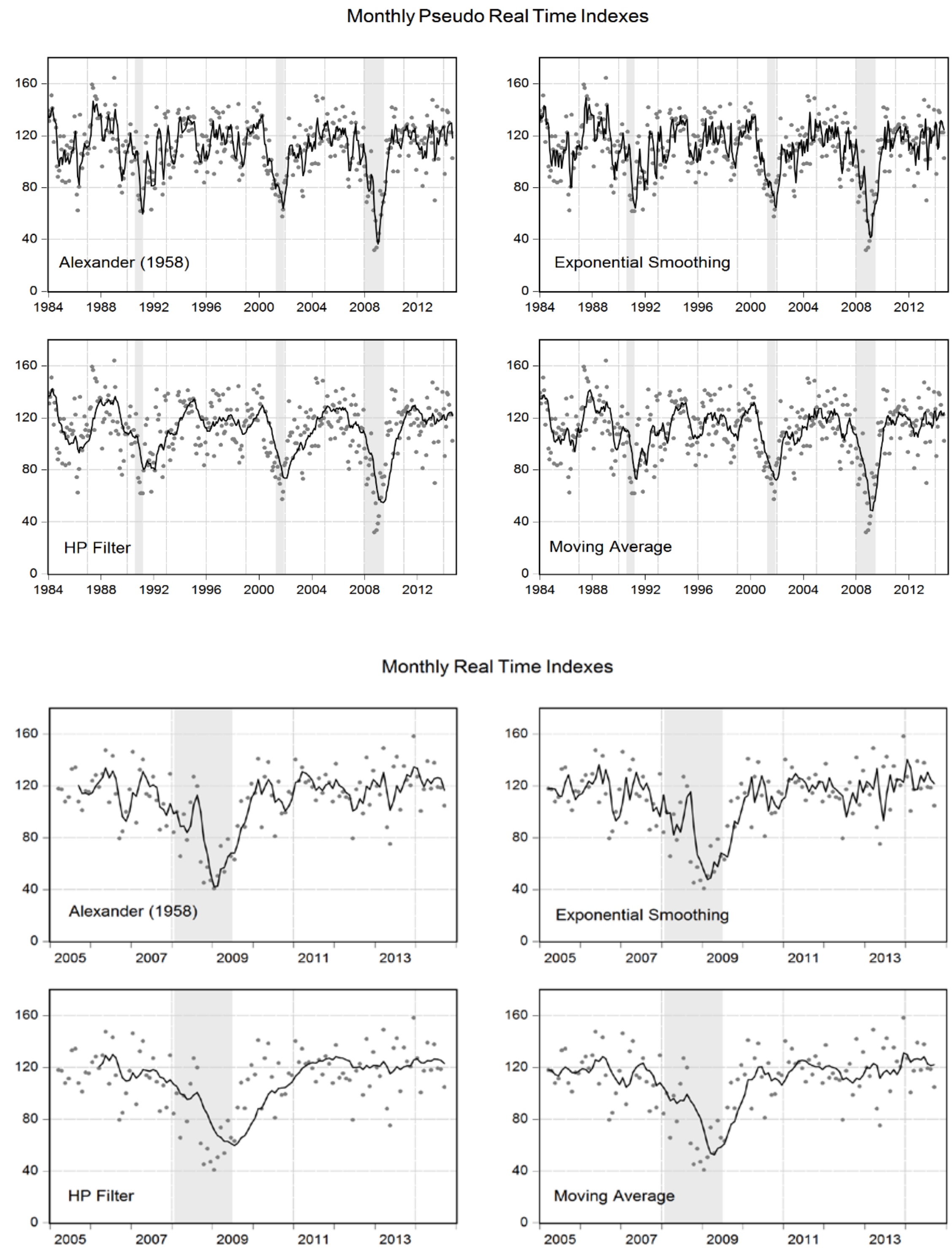

Figure 2. Signals from Diffusion Index, Various Rules, Pseudo- and Real-Time, Dates of Cycles

This figure shows binary nowcasts of recession made using alternative decision rules. The lightly shaded areas are recession periods. Areas with dark shades are predicted recession periods. Note that to facilitate comparison, dark shades are intentionally drawn shorter than light shades. The label “3 declines” refers to the three consecutive decline rule.

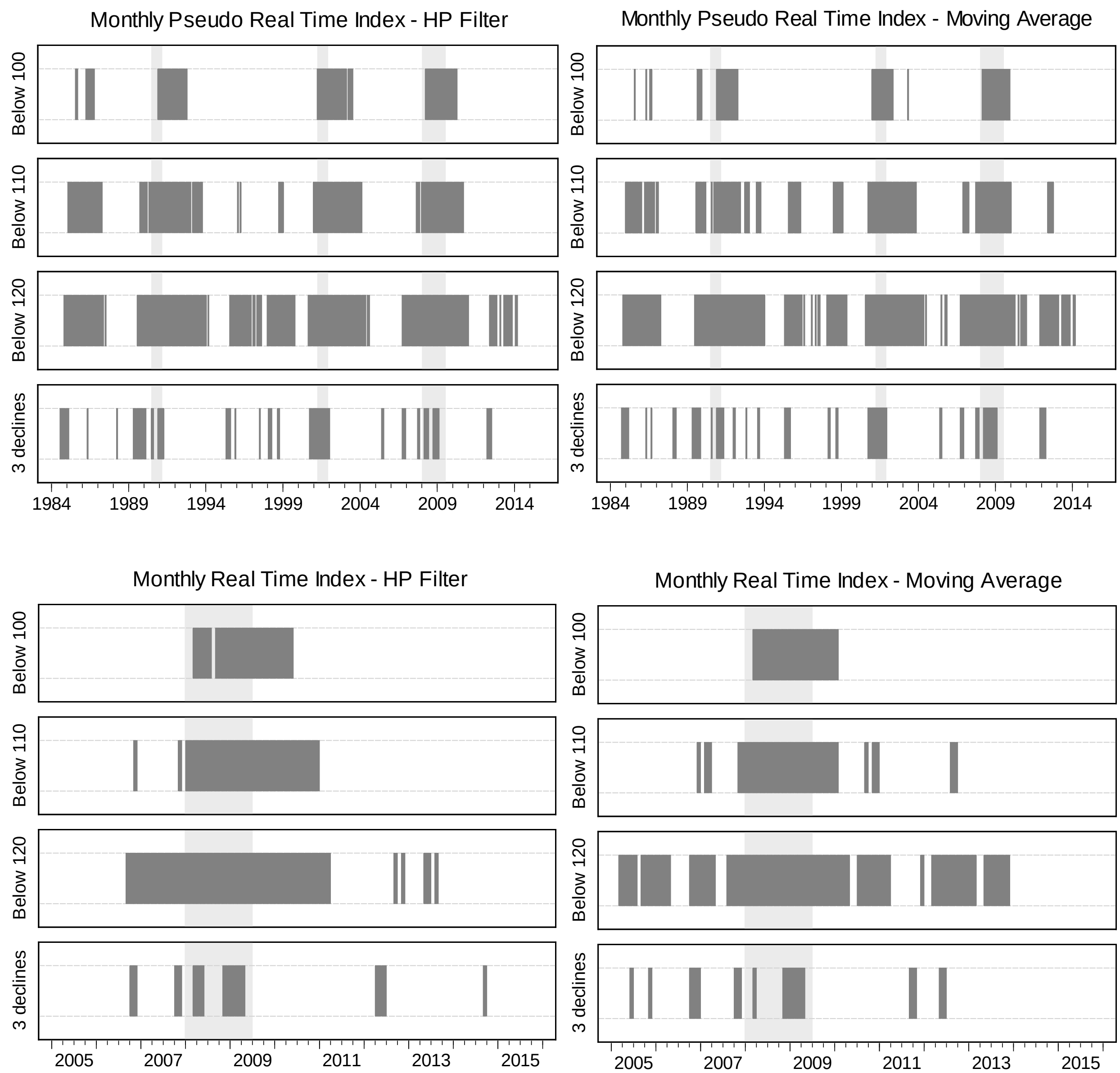

Figure 3. Probability of a Recession, Obtained from Probit with three Lags

This figure shows probability forecasts made using probit models with three lags of the smoothed index (moving average smoothing) on the right-hand-side.