BUSINESS ANALYTICS WORKING PAPER SERIES

Forecast combination for U.S. recessions with real-time data

Andrey L. Vasnev and Laurent L. Pauwels
University of Sydney

Abstract

This paper proposes the use of forecast combination to improve predictive accuracy in forecasting the U.S. business cycle index as published by the Business Cycle Dating Committee of the NBER. It focuses on one-step ahead out-of-sample monthly forecast utilising the well-established coincident indicators and yield curve models, allowing for dynamics and real-time data revisions. Forecast combinations use log-score and quadratic-score based weights, which change over time. This paper finds that forecast accuracy improves when combining the probability forecasts of both the coincident indicators model and the yield curve model, compared to each model's own forecasting performance.

January 2013

BA Working Paper No: 02/2013

http://sydney.edu.au/business/business_analytics/research/working_papers
Forecast combination for U.S. recessions with real-time data

Laurent Pauwels,∗,∗∗, Andrey Vasnev

aUniversity of Sydney, New South Wales 2006, Australia

Abstract
This paper proposes the use of forecast combination to improve predictive accuracy in forecasting the U.S. business cycle index as published by the Business Cycle Dating Committee of the NBER. It focuses on one-step ahead out-of-sample monthly forecast utilising the well-established coincident indicators and yield curve models, allowing for dynamics and real-time data revisions. Forecast combinations use log-score and quadratic-score based weights, which change over time. This paper finds that forecast accuracy improves when combining the probability forecasts of both the coincident indicators model and the yield curve model, compared to each model’s own forecasting performance.

Keywords: U.S. business cycle, Forecast combination, Density forecast, Probit models, Yield curve, Coincident indicators.

JEL: C5, C3

1. Introduction

Macroeconomic research modelling U.S. economic conditions divides the business cycle into two distinct states: periods of economic growth, or expansions, and periods of economic contraction, or recessions. Modelling and forecasting the U.S. business cycle is still very much topical in macroeconomic research as seen in the recent Econometrica publication by Schmitt-Grohé and Uribe (2012). The Business Cycle Dating Committee of the NBER defines the period from a peak to a trough as a recession, while an expansion is the period extending from a trough to a peak. These published peaks and trough periods can be used to construct a binary recession index.

The NBER-dated binary recession indicator lends itself naturally to a probit model. The underlying state of the economy can be modelled as

\[ Y_{t+1}^* = \mathbf{x}_t' \beta + \varepsilon_{t+1} | \mathbf{x}_t \sim \text{i.i.d. } N(0, 1) \]  

where \( Y_t^* \) is the unobserved latent variable and \( \mathbf{x}_t \) is the \((1 \times k)\) vector of explanatory variables with the corresponding coefficient vector \( \beta \). The NBER recession index \( Y_t \) is observed such that

\[ Y_{t+1} = \begin{cases} 
0 & \text{if } Y_{t+1}^* \geq 0 \\
1 & \text{if } Y_{t+1}^* < 0 
\end{cases} \]

Typically, the literature uses either of two sets of covariates to model the underlying economic conditions in the U.S. (1) The four coincident indicators: real manufacturing and retail trade sales (sales), total personal income less transfer payments (income), the civilian labour force employed in non-agricultural industries (employment), and industrial production (IP); and (2) the yield curve. The four coincident indicators remain listed as the key decision variables used by the NBER’s Business Cycle Dating Committee.

The yield curve is defined as the spread between the 10 year treasury bond rate and the 3 month bill rate. It is considered to be a leading indicator of economic activity and an alternative to the coincident indicators model, see Chauvet and Potter (2002) and Stock and Watson (2003). Kauppi and Saikkonen (2008) asserts that the yield curve is the single best out-of-sample predictor for U.S. recessions.

For either model with coincident indicators or the yield curve, equation (1) can be augmented to capture the persistence in the business cycle by lagging the recession index, \( Y_t \) as follows

\[
Y^*_t = x'_t \beta + \theta Y_t + \varepsilon_{t+1} | x_t, Y_t \sim \text{i.i.d.} \ N(0, 1)
\]

where \( \theta \) is the autoregressive parameter (\( |\theta| < 1 \)). The main advantage to including \( Y_t \) is to account for serial correlation which manifests itself through high degree of persistence and dependence in the occurrence of recessions and expansions (see Chauvet and Potter, 2005). However, one limitation of including the NBER recession index is that it is published with substantial delay and thus the models using the lag do not reflect real-time forecasting conditions.

The focus of this paper is to evaluate the out-of-sample forecasting performance of the combined recession probability forecasts of the coincident indicators and the yield curve models. This is compared to the forecasting performance of these two models which are so frequently used in the literature. Diverse combination schemes are also investigated. This paper uses scoring rules as a way to evaluate the forecasting performance of the models. The rest of the paper is organised as follows. Section 2 discusses the forecasting combination methodology and the data used in the paper is presented in section 3. Section 4 discusses the empirical results and section 5 concludes.

2. Methodology

We start with two competing models to forecast the U.S. recessions. Rather than simply identifying which of these provides superior forecasting performance, we apply forecast combination techniques as a way to improve forecast accuracy (see Timmermann, 2006 for a survey), robustness against structural breaks, model misspecification and measurement errors (Stock and Watson, 2001, 2004). We combine one-step ahead recession probability forecasts of the two competing probit models. The recession and expansion probability
forecasts can be combined in a \((2 \times 1)\) vector

\[
\hat{P}_{i,t+1|t} = \left( \hat{P}_{i,0,t+1|t}, \hat{P}_{i,1,t+1|t} \right),
\]

where \(\hat{P}_{i,0,t+1|t}\) is the probability of an expansion and \(\hat{P}_{i,1,t+1|t}\) is the probability of a recession for model \(i\). A simple way of combining probability forecast vectors from different models is

\[
\hat{P}_{W,t+1|t} = \sum_{i=1}^{n} \alpha_i \hat{P}_{i,t+1|t}
\]

where \(\sum_{i=1}^{n} \alpha_i = 1\) and \(n\) is the number of models combined and \(W\) is an index for the method of weighting.

The weighting schemes considered here follow the methodology developed by Pauwels and Vasnev (2011). We consider equal weights, where \(\alpha_i = 1/n\), and two types of adaptive weights. The adaptive weights are constructed from average scores. The scoring rules for each period are given by

\[
S^L_{i,t+1|t} = \log(\hat{P}_{i,j,t+1|t}),
\]

\[
S^Q_{i,t+1|t} = 2\hat{P}_{i,j,t+1|t} - \left[ (\hat{P}_{i,0,t+1|t})^2 + (\hat{P}_{i,1,t+1|t})^2 \right],
\]

where \(S^L_{i,t+1|t}\) and \(S^Q_{i,t+1|t}\) are the log and quadratic scores for the \(i\)th model at time period \(t + 1\). The actual observed state is given by \(j\), where \(j = 0, 1\). The scores aim to maximise the sharpness of the predictive distribution, and hence more accurate models are assigned a higher weight through higher log and quadratic score. We also use the scores as a summary measure of the predictive performance, thus enabling easy comparison and ranking of all of the model specifications under consideration.

\[3. \text{Data}\]

The time frame that is used in this paper is limited by the availability of the explanatory variables and spans from January 1967 to June 2010.\(^1\) The explanatory variables are released on a monthly basis and are calculated as year-on-year growth rate (except for the yield curve). We do not assume full information at the forecast origin, rather we limit the data used in estimation to what would have actually been known. There are two elements to this. First, we use the data from the month \(t\) to estimate out-of-sample one-step ahead forecasts for \(Y_{t+1}^*\). Second, as values of macroeconomic indicators are regularly revised, their most recent values assume knowledge of future data revisions.

In order to overcome this latter weakness, we use real-time data available for some of the indices. Indeed, the signals sent from the real-time data are often different to the image that emerges after the revisions have taken place, especially when business cycles are at turning

\(^1\)Available at http://www.nber.org/cycles/cyclesmain.html, the data for this paper was downloaded in March 2011.
points (see Hamilton, 2010). The Federal Reserve Bank of Philadelphia are pioneers in the construction of real-time data series for the U.S.\textsuperscript{2} Real-time data matrices are depicted in Table 1 for example. For the variable $x$ we will indicate the value of $x_t$ at vintage $s$ by $x(t, s)$. Typically, the observation for month $t$ is initially released in month $t + 1$ and thus first appears in vintage $t + 1$. It is then updated in future vintages. The changes will decrease in magnitude, and should reach zero with a delay of six months to two years. Instead of moving downwards as time progresses, we move diagonally with an extra observation and vintage available with each step forward we make. Thus, not only do we have access to an additional observation but the most recent past observations are updated.

Table 1: Illustration of a real-time data matrix

\begin{tabular}{|c|c|c|c|c|}
\hline
\text{← vintages} & January & February & March & April & May \\
\hline
\text{↑} & January & - & $x(1, 2)$ & $x(1, 3)$ & $x(1, 4)$ & $x(1, 5)$ \\
\hline
\text{time} & February & - & - & $x(2, 3)$ & $x(2, 4)$ & $x(2, 5)$ \\
\hline
\text{↓} & March & - & - & - & $x(3, 4)$ & $x(3, 5)$ \\
\hline
\text{↓} & April & - & - & - & - & $x(4, 5)$ \\
\hline
\end{tabular}

Several papers compare the forecasts obtained when using final vintage versus real-time data. Chauvet and Hamilton (2006) finds model estimation with real-time data reduces the quality of estimates, due to the additional noise compared to latest-vintage data. Chauvet and Piger (2008) oppose this conclusion in their application of real-time data in a Markov-switching dynamic factor model to business cycle turning points, concluding that data revisions do not appear to significantly effect the estimated business cycle turning points (see Hamilton, 1989 and Chauvet, 1998 for details on this model).

Of the four coincident indicators, only non-farm payroll employment and the index of industrial production are publicly available as real-time data. Real manufacturing and trade sales and real personal income excluding transfer payments are yet to be constructed by the Philadelphia Fed. While Chauvet and Potter (2005), Chauvet and Hamilton (2006) and Chauvet and Piger (2008) created these variables in real-time, these series have not be made public. Hence, the latest vintage values of sales and income will be used.

The real-time data series are released with a delay of one month, that is, the initial value for month $t$ is released in month $t + 1$ and is updated in subsequent months. We follow Chauvet and Piger (2008) in timing the variables with latest vintage data in the same model. Hence, to estimate our forecast for month $t + 1$ at the forecast origin $t$, we use the observation of all four coincident indicators at month $t - 1$.

In Figure 1 we juxtapose the year-on-year growth rates for the latest vintage and real-time data series of non-farm payroll employment and industrial production, for which we have real-time data.

Figure 1: Comparison of the latest vintage vs. real time year-on-year growth rates of employment and industrial production

Two remarks are noteworthy with regards to Figure 1. First, the real-time values of year-on-year growth in employment and in IP are lagging the latest vintage estimates. Furthermore, the variance is higher for real-time data than latest vintage, a result of the inherent additional uncertainty. The difference between the two series is especially pronounced at the turning points. This is integral as these are the periods when we most need reliable estimates of the underlying state of the economy to determine the turning points of the business cycle. Second, the difference in the real-time and latest vintage series lessens as we reach the end of our sample. In part, this could be explained by the proximity of these values to the date of collection (March 2011), and hence the latest vintage values are still undergoing revision.

The yield curve is constructed as the difference in the interest rates of long- and short-term bonds. Following Chauvet and Potter (2002, 2005), Kauppi and Saikkonen (2008) and Kauppi (2010) these are the 10-year Treasury bond rate and three-month Treasury bill rate. We use the 10-year Treasury constant Maturity rate, released monthly and the three-month Treasury bill secondary market rate, similarly released on a monthly basis. Given that the yield curve is strictly comprised of Treasury bond and bill rates, these are naturally never revised and hence there is no real-time data dimension to this model. Finally, all series have been tested for stationarity using ADF unit root tests. The null of hypothesis of a unit root is rejected for all series (results available upon request).

---

3 This is accentuated by the NBER-published data not containing revisions to March 2011, but rather to the 18th September 2010. Given this date is three months after our sample ends, we suspect the latest vintage values from early 2009 onwards to still require some degree of adjustment.

4. Forecasting U.S. recessions

We present results for out-of-sample probability forecast using an expanding window. Forecasting is conducted as follows. Initial values of the model parameters are estimated using the first half of the sample, with observations $t = 1, \ldots, T/2$, and then the parameter estimates are recursively updated. The forecasted periods span from March 1989 until June 2010. Note that when we work with real-time data, we not only obtain an additional observation when recursively estimating the log-likelihood function but a revised sample, as previous values of the real-time variables are updated. As a robustness check, we also considered rolling window estimation by setting a window size of half the sample, which is not presented here. The results are available upon request.

We conduct this empirical experiment for several model specifications. First, we provide results for the two benchmark models, the coincident indicators model and the yield curve model, plus a model featuring both coincident indicators and the yield curve as covariates (“Coincident var. + Yield var.”). These three models are also re-estimated with the inclusion of the lagged recession indicator. Second, using forecast combination techniques as described in the earlier section, we combine the probability forecasts of the coincident indicators model and the dynamic yield curve model. Similarly to Kauppi and Saikkonen (2008) and Kauppi (2010), we find that the performance of the yield curve model can be improved through the inclusion of a lagged recession indicator. Hence, we use the dynamic rather than static yield curve model recession probabilities. Lastly, we also present results for combination of probability forecasts for five univariate models, composed of one of the four coincident indicators or the yield curve. In the dynamic case, the univariate models contain a lagged recession indicator.

4.1. Results

The success/failure matrix in Table 2 shows that the inclusion of the lagged recession indicator (“Dynamic”) consistently leads to better overall prediction compared to the “static” specification. Moreover, the joint model grouping the coincident indicators, the yield curve and the lagged recession indicator (“Coincident var. + Yield var.”) produces better overall prediction than either model on their own. More importantly, Table 3 provides evidence that a simple combination, which weights the two models' forecast probabilities equally, outperforms the three models presented in Table 2.
Table 2: Success/Failure matrices for three multivariable models

<table>
<thead>
<tr>
<th></th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yield curve</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>223</td>
<td>218</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>100.00</td>
<td>97.76</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>85.77</td>
<td>90.77</td>
</tr>
<tr>
<td><strong>Coincident indicators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>212</td>
<td>212</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>97.25</td>
<td>97.25</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>89.02</td>
<td>89.41</td>
</tr>
<tr>
<td><strong>Coincident var. + Yield var.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>212</td>
<td>206</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>97.25</td>
<td>94.50</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>89.41</td>
<td>92.55</td>
</tr>
</tbody>
</table>

Notes: The three models are the coincident indicators model, the yield curve model and a model combining regressors: the coincident indicators and the yield curve (Coincident var. + Yield var.). The forecasting period spans from March 1989 until June 2010.

Table 3: Success/Failure matrices for forecast combination of the coincident indicators and the yield curve models

<table>
<thead>
<tr>
<th>Combination method</th>
<th>Static</th>
<th>Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equal weights</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>214</td>
<td>212</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>98.17</td>
<td>97.70</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>95.69</td>
<td>95.28</td>
</tr>
<tr>
<td><strong>Log score weight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>28</td>
<td>212</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>97.70</td>
<td>81.08</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>95.28</td>
<td></td>
</tr>
<tr>
<td><strong>Quad score weight</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expansion (predicted)</td>
<td>212</td>
<td>212</td>
</tr>
<tr>
<td>Recession (predicted)</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Correct prediction (%)</td>
<td>97.70</td>
<td>81.08</td>
</tr>
<tr>
<td>Overall correct prediction (%)</td>
<td>95.28</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Predictions when the coincident indicators and the yield curve models are combined across different weighting schemes. The forecasting period spans from March 1989 until June 2010.
The forecast combination model of the coincident indicators and the yield curve models with log score weight outperforms all other models both in terms of log and quadratic scoring rules as shown in Table 4. This evidence corroborates the results found in the success/failure matrices of Tables 2 and 3. Note also that the dynamic coincident model’s performance is comparable to the forecast combination model of the coincident indicators and yield curve models in terms of quadratic scoring rule. However, this is not the case when looking at log-scores. Log-scoring rules provide a natural theoretical justification for evaluating density forecasts. It is closely related to the Kullback-Leibler Information Criterion (KLIC) distance, as KLIC is the expectation of the log-densities (see Hall and Mitchell, 2007).

In the forecast combinations of the recession probabilities of the coincident indicators model and yield curve model, the yield curve model has higher weight at the beginning of the forecasting sample. This weight reduces from approximately 80% to 50% as a consequence of poor performance during the 1990 recession. After this recession, each model is assigned approximately 50% weight. The weight assigned to the coincident indicator model gradually falls away as the yield curve model performs accurately in periods of expansion, notably during the large gap between the 1990 and 2001 recessions. During the 2008 recession (double the length of the previous two recessions) some of the weight is reallocated to the coincident indicators model.

Figure 2: The behaviour of the Log weights over time

Notes: Combination of predictions from the coincident indicators and the yield curve models using a log weighting scheme. The forecasting period spans from March 1989 until June 2010.
Table 4: Forecast combination model score functions

<table>
<thead>
<tr>
<th>Model</th>
<th>Benchmark models</th>
<th>Log</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Static</td>
<td>-0.424</td>
<td>0.750</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.753</td>
<td>0.831</td>
</tr>
<tr>
<td>Coincident indicators</td>
<td>Static</td>
<td>-0.253</td>
<td>0.850</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.427</td>
<td>0.908</td>
</tr>
<tr>
<td>Coincident var. + Yield var.</td>
<td>Static</td>
<td>-0.244</td>
<td>0.853</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.644</td>
<td>0.854</td>
</tr>
<tr>
<td>Forecast combinations of Coincident &amp; Yield models</td>
<td>Log score weighted</td>
<td>-0.180</td>
<td>0.909</td>
</tr>
<tr>
<td>Forecast combinations of 5 univariate models</td>
<td>Quadratic score weighted</td>
<td>-0.182</td>
<td>0.907</td>
</tr>
<tr>
<td>Equal weights</td>
<td>Static</td>
<td>-0.328</td>
<td>0.811</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.319</td>
<td>0.814</td>
</tr>
<tr>
<td>Log score weighted</td>
<td>Static</td>
<td>-0.322</td>
<td>0.815</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.339</td>
<td>0.817</td>
</tr>
<tr>
<td>Quadratic score weighted</td>
<td>Static</td>
<td>-0.327</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td>-0.318</td>
<td>0.814</td>
</tr>
</tbody>
</table>

Note: Dynamic refers to the inclusion of a lagged recession indicator in the models or as a separate model in the univariate model forecast combination. A dynamic yield curve model is combined with the static coincident indicators model in the “Coincident & Yield models” results. The 5 univariate models are composed of one of the four coincident indicators or the yield curve. In the dynamic case, the univariate models contain a lagged recession indicator.

Figure 3 shows that the dynamic specification of the three benchmark models result in recession probabilities that are sharper than those generated by the static version of these models in that they are concentrated towards zero and one. When comparing Figures 3 and 4, one can discern that forecast combination acts to moderate the forecasts, they become less sharp.
Figure 3: Recession probabilities for benchmark models

(a) Static yield curve
(b) Dynamic yield curve
(c) Static coincident indicators
(d) Dynamic coincident indicators
(e) Coincident var. + Yield var. (Static)
(f) Coincident var. + Yield var. (Dynamic)
Figure 4: Recession probabilities when predictions of the coincident indicators and the Dynamic yield curve models are combined with three weighting schemes.

5. Conclusion

This paper examines the out-of-sample forecast performance of the well-established coincident indicators and yield curve models, allowing real-time revisions to the employment and industrial production data. This paper finds that forecast combination of both the coincident indicators model and yield curve model improves forecast accuracy compared to each of the models’ own forecasting performance. Furthermore, the empirical evidence is in favour of model combination rather than combining the regressors into one model.

Acknowledgements

The authors are grateful for Emma Smith’s research assistance. The University of Sydney Business School’s Research Grant 2012 is acknowledged.

References


