Real Time Uncertainty in Business Cycle Dating

Domenico Giannone (Université Libre de Bruxelles)

September 21, 2012

Let y_t be the log-of GDP. We are interested in identifying the dates of peaks and trough. Defined as follows:

Definition Peak: I say that a peak occurred in quarter t if

$$y_{t-2}, y_{t-1} \leq y_t \geq y_{t+1}, y_{t+2}$$

Definition Trough: I say that a trough occurred in quarter t if

$$y_{t-2}, y_{t-1} \geq y_t \leq y_{t+1}, y_{t+2}$$

I will also impose that there must be at least one quarter distance between Peak and Trough.

The definition used here is a simplified version of the Bry and Boshan dating algorithm. Research has shown that when applied to historical data the algorithm identifies roughly the same dates of the NBER dating committee, see Harding and Pagan, 2002.

Remark Journalistic rule that dates a peak if there are two consecutive quarters of negative growth is more stringent than the rule derived above.

Remark The dating rule is usually applied to the level, rather than log-level, of the series. I work on log-levels for simplicity. Results are unaffected by this choice.

When applied to the Euro Area data from 1999 the algorithm identifies the same dates as the CEPR committee: peak in 2008q1 followed by a trough in 2009q2. These dates are compatible with those identified by the CEPR dating committee. In addition the algorithm identifies a peak in 2011q3.

The algorithm did not date an early millennium recession in Europe. The reason is that in 2001q1 there was a major decline but the economy recovered after one quarter. In addition, with real time data also the decline disappeared (see Giannone et al., 2012).

As I said above, the algorithm also identifies a peak in 2011q3. The data are going to be subsequently revised especially those for the most recent observation and the revisions can be such that the peak will disappear in subsequent vintages. The aim of this note is to provide an assessment of the probability that a peak will not be identified when looking at the data three years from now. The question is: if today we declare a peak, what is the probability that when we will meet in September 2015 we will realize that this was a mistake?

I will try to answer this question by using the historical correlations among different revisions in order to forecast the data that will be available in the September 2015.

Table 1: Probability that we will detect a peak in a given quarter with the vintage of data of September 2015

P(peak at t)	2010				2011				2012	
(%)	q1	q2	q3	q4	q1	q2	q3	q4	q1	q2
Jun-2012	0	0	0	0	9	18	64	0	0	0
Sep-2012	0	0	0	0	4	18	**74**	1	0	0

The table report the probability that the data that will be available in three years will show a peak in a specific quarter. The exercise is computed using the data available with the Monthly Bulletin of June 2012, which contained information for GDP up to 2012q1, and with the Monthly Bulletin of September 2012 (which contains the first release for GP growth in 2012q2).

Given that the events of peak in specific quarters are disjoint, the probability that data revisions would undo the evidence about the occurrence of a peak in 2011 was about 10% in June. The probability that the peak in 2012q3 would be confirmed was 64%.

The data that has been released during the summer has strengthen this message and now the probability of wrongly dating a recessions in 2011 has reduced to 5%. Also the probability that the peak in 2012q3 will be confirmed is around 75%

Details of the methodology and implementations are provided below.

0.1 Additional remarks

The exercise described above just focuses on GDP. However, similar results hold for the other data series taken individually. Please let me know if you prefer to have the details for the analysis of the other variables.

We could also consider a joint model for all variables. If we do so, we should expect the probability that the currently observed peak is due to measurement error to become even smaller. The reason is that the joint event of getting two errors in the same direction, say for GDP and employment, is smaller or equal than the probability of making the mistake in each variables separately. In this sense 5% should be understood as an upper bound on the probabilities of wrongly dating the recession in 2011.

1 Methodology

1.1 Data

Define $\Delta y_{v,t}$ the quarterly growth rate of real GDP in quarter t as available in vintage v.

I use the Euro Area Real-Time database constructed by Giannone et al. (2012). The database has been constructed by using the vintages of the ECB Monthly Bulletin. This means that the frequency of the vintages (monthly) is potentially higher than the frequency of the data (quarterly). Accounting for this difference in frequency of data and vintages can make the empirical analysis unnecessarily complicated. To avoid this problem we will focus on quarterly vintages corresponding with the Monthly Bulletin issued of the third month of each quarter. Fixing $t = 12q2$, $\Delta y_{t+i,t}$ is GDP growth for $12q2$ as it will be reported in different subsequent vintages of the Monthly Bulletin. For $i = 1$ we have GDP growth for $12q2$ as recorded in the ECB Monthly Bulletin of September 2012.[1] This is the first time GDP growth for $12q2$ appears on the Monthly Bulletin.[2] The data will correspond to the first estimate of Eurostat. For $i = 4$ we have GDP growth for $12q2$ as it will be recorded in the ECB Monthly Bulletin of September 2013. For $i = 12$ we have GDP growth for $12q2$ as it will be recorded in the ECB Monthly Bulletin of September 2015.

1.2 Model

$$\Delta y_{t+i,t} = \mu_i + \lambda_{i,0} f_t + \lambda_{i,1} f_{t-1} + \rho_{i,1} \Delta y_{t+i,t-1} + e_{t+i,t}, \quad i = 1, ...36$$

The noise $e_{i,t}$ is assumed to have mean zero and not to be correlated neither across vintages i nor across time $\mathrm{E}(e_{i,t}) = 0$ and $\mathrm{E}(e_{i,t} e_{j,t'}) = 0$ for all $i \neq j$ and $t \neq t'$. The common factor f_t is assumed to follow an Autoregressive process of order 1.

Remark: The model above nests the noise model of data revisions according to which each release is equal to a true underlying concept plus a noise: $\Delta y_{i,t} = \Delta y_t^* + e_{i,t}$.

The model is clearly miss-specified since the assumption because of the exact factor structure that assumes that the errors for the same quarter but associated with subsequent vintages are uncorrelated, $\mathrm{E}(e_{i,t} e_{j,t}) = 0$, $i \neq j$. With some more work I can modify the model to assume more realistic data generating process. However, fortunately the qualitative results would be confirmed since the estimates of the model used here have been shown to be robust to such form of miss-specifications (see Doz et al., 2012).

[1] As a rule, the cut-off date for statistics included in the Monthly Bulletin data is scheduled for the day before the first Governing Council meeting in a month. The cut-off date for the statistics included in the September 2012 issue is 5 September 2012.

[2] The flash estimate for GDP growth is released 6 weeks after the end of the quarter. The first release is published 8 weeks after the end of the reference quarter. As an example, for $12q2$ the flash estimate was published ion August 14 and the first release on August ??.

1.3 Assessing Real-Time Uncertainty in Real-Time Data-Releases

I estimate the model using vintages from $02q2, ..., 12q3$ and considering revisions up to 3 years after the reference quarter ($i = 1, ..., 12$).

We use a rather diffuse bayesian prior I draw 1000 path from the posterior probabilities.

We draw from the posterior probabilities of the values of $\Delta y_{i,t}$ that are not available. These are posterior probabilities in the sense that they are based on all the information available. Technically, these are conditional forecasts since the available information in the vector of observations are not aligned. Draws from predictive densities in this case can be easily computed using the Kalman filter (see Banbura, Giannone and Lenza, 2012).

When looking at the posterior I account for both uncertainty in the model, and uncertainty in the shocks hitting the model.

For each i I draw a 1000 paths for the missing data. For any given path I identify the peaks using the rule defined above on $y_{13q3,t}$. Log-levels are recovered by cumulating the growth rates. This tells us what is the probability that when I will apply the dating rule on revised data after three years I will identify a peak. This probability is computed as the percentage of the occurrence that a peak occurs at any given point in time.

References

Catherine Doz & Domenico Giannone & Lucrezia Reichlin, 2006. "A quasi maximum likelihood approach for large approximate dynamic factor models," Working Paper Series 674, European Central Bank. Review of Economics and Statistics, forthcoming.

Domenico Giannone & Jerome Henry & Magdalena Lalik & Michele Modugno, 2010. "An area-wide real-time database for the euro area," Working Paper Series 1145, European Central Bank. Review of Economics and Statistics, forthcoming.

Harding, Don & Pagan, Adrian, 2002. "Dissecting the cycle: a methodological investigation," Journal of Monetary Economics, Elsevier, vol. 49(2), pages 365-381, March.