Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 333–346
international journal of forecasting
www.elsevier.com/locate/ijforecast

A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP

Marta Bańbura[a,*], Gerhard Rünstler[b]

[a] *European Central Bank, Germany*
[b] *Austrian Institute of Economic Research, Austria*

Available online 15 April 2010

Abstract

We derive forecast weights and uncertainty measures for assessing the roles of individual series in a dynamic factor model (DFM) for forecasting the euro area GDP from monthly indicators. The use of the Kalman smoother allows us to deal with publication lags when calculating the above measures. We find that surveys and financial data contain important information for the GDP forecasts beyond the monthly real activity measures. However, this is discovered only if their more timely publication is taken into account properly. Differences in publication lags play a very important role and should be considered in forecast evaluation.

Keywords: Dynamic factor models; Filter weights; GDP; Publication lags

1. Introduction

The first estimate of euro area quarterly GDP is released about six weeks after the end of the quarter. For assessing macroeconomic conditions in the meantime, forecasters rely on data of higher frequencies, including financial series, surveys, and monthly data on real economic activity. The first two categories reflect market expectations and are often labelled as 'soft' data, as opposed to the 'hard' indicators of real activity that measure certain components of the GDP directly (e.g. industrial production). The soft data are available promptly, while real activity data are published with a significant delay. On the other hand, the latter category is considered to contain a more precise signal for GDP. Overall, the large number of indicators available, with their different release dates and potential trade-offs between timeliness and quality, makes it difficult to use the information contained in the various indicators efficiently. Put differently, it is not straightforward to

* Corresponding author.
E-mail addresses: marta.banbura@ecb.europa.eu, marta.banbura@ecb.int (M. Bańbura), gerhard.ruenstler@wifo.ac.at (G. Rünstler).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2010.01.011

attach appropriate weights to the individual indicators when producing a GDP forecast.

Factor models have emerged as an interesting alternative for the short-term forecasting of real activity, as they allow information to be extracted from large data sets in a parsimonious manner.[1] Recently, Giannone, Reichlin, and Small (2008) proposed a factor model framework that can cope with large data sets which are characterised by staggered and nonsynchronous data releases, resulting in an irregular pattern of missing data at the end of the sample (sometimes referred to as "ragged edge" data). This framework implements the common factors as unobserved components in a state space form. The Kalman smoother provides efficient forecasts, also in situations where there are missing data points in past observations. As a result, new information can be incorporated as soon as it arrives. The weight of this information in the forecast is attached in an automatic manner, taking into account both the timeliness and the quality of the signal. Giannone et al. (2008) apply this model to produce projections of the US GDP from a large number of 'hard' and 'soft' monthly indicators. They show that the forecast accuracy increases monotonically with each release; hence, it is important to incorporate the latest information.

In this paper we build on the framework of Giannone et al. (2008) and implement it for the euro area.[2] The main purpose of our paper is to investigate the role of real activity, survey and financial data in forecasting euro area GDP. In particular, we discuss a method of deriving the weights of individual series in the forecast. We investigate how these forecast weights evolve with the advent of new information and the way in which they depend on the timeliness of different indicators. In addition, we consider uncertainty measures similar to those employed by Giannone et al. (2008), to assess the marginal contributions of different groups to the forecast accuracy.

In general, studies report that soft data contain little information beyond the real activity data; however, they mostly ignore differences in timeliness (e.g. Banerjee, Marcellino, & Masten, 2005; Forni et al., 2003; Stock & Watson, 2003). Giannone et al. (2008) construct pseudo real-time data sets that follow the actual data availability closely, and evaluate the impact of different data releases on forecast precision. They find that, after properly accounting for the more timely availability of surveys relative to real activity data (e.g. industrial production), the largest gains in forecast precision occur after the release of surveys.[3]

In this paper we study this issue in more detail for euro area data. Since hard indicators have longer publication delays in the euro area than in the US, there may be an even larger role for soft data. Following Giannone et al. (2008) and Rünstler and Sédillot (2003), we construct euro area pseudo real-time data sets that replicate the data availability within each month of the sample.

We extend the work of Giannone et al. (2008) in a number of directions. First, while Giannone et al. (2008) used uncertainty measures to assess the role of new data releases, we also derive the weights of the series in the forecast. We show how a method proposed by Koopman and Harvey (2003) can be used to calculate forecast weights in a real-time context, i.e. with non-synchronous data releases. We regard forecast weights as a complementary tool to the uncertainty measures. While the latter give the marginal gain in forecast precision from a certain data release, weights show the actual contribution of individual series to a forecast. Therefore, weights are also helpful in understanding the role of individual data in specific forecast episodes and forecast revisions.

Second, to investigate the relative roles of timeliness and quality we repeat the exercises under counterfactual assumptions. For our main exercise, we use a pseudo real-time design to replicate the information sets that are available in each month within the quarter (Giannone et al., 2008; Rünstler & Sédillot, 2003). However, we also construct two counterfactual data

[1] See e.g. Artis, Banerjee, and Marcellino (2005), Bernanke and Boivin (2003), Boivin and Ng (2005), D'Agostino and Giannone (2006), Forni, Hallin, Lippi, and Reichlin (2003, 2005), Giannone, Reichlin, and Sala (2004), Marcellino, Stock, and Watson (2003) and Stock and Watson (2002a,b).

[2] While our paper is the first application of the model by Giannone et al. (2008) to the euro area, further applications were conducted by Angelini, Bańbura, and Rünstler (2008), Angelini, Camba-Méndez, Giannone, Rünstler, and Reichlin (2008). For applications to other economies, see e.g. Aastveit and Trovik (2007) and Matheson (2010). See e.g. Schumacher and Breitung (2008) for an application of the framework of Stock and Watson (2002b) in a real-time data flow context.

[3] Hansson, Jansson, and Löf (2005) also report that the inclusion of survey data in VAR models improves out-of-sample forecasts; however, they use a small data set and quarterly data.