Essays in Dynamic Macroeconometrics

Key words: Large cross-section, Factor model, EM algorithm, Bayesian VAR, Wavelets

Summary

The thesis contains four essays covering topics in the field of macroeconomic forecasting. The first two chapters consider factor models in the context of real-time forecasting with many indicators. Using a large number of predictors offers an opportunity to exploit a rich information set and is also considered to be a more robust approach in the presence of instabilities. On the other hand, it poses a challenge of how to extract the relevant information in a parsimonious way. Recent research shows that factor models provide an answer to this problem.\(^1\) The fundamental assumption underlying those models is that most of the co-movement of the variables in a given dataset can be summarized by only few latent variables, the factors. This assumption seems to be warranted in the case of macroeconomic and financial data. Important theoretical foundations for large factor models were laid by Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002a). Since then, different versions of factor models have been applied for forecasting\(^2\), structural analysis\(^3\) or construction of economic activity indicators\(^4\). Recently, Giannone, Reichlin, and Small (2008) have used a factor model to produce projections of the U.S. GDP in the presence of a real-time data flow. They propose a framework that can cope with large datasets characterised by staggered and nonsynchronous data releases (sometimes referred to as “ragged edge”). This is relevant as, in practice, important indicators like GDP are released with a substantial delay and in the meantime more timely variables can be used to assess the current state of the economy.

The chapter A look into the factor model black box: publication lags and the role of hard and soft data in forecasting GDP is based on a joint work with Gerhard Rünstler and applies the framework of Giannone, Reichlin, and Small (2008) to the case of euro area. In particular, we are interested in the role of “soft” and “hard” data in the GDP forecast

\(^1\)Alternative solutions include e.g. forecast combination or Bayesian model averaging, see Stock and Watson (2006) for a survey.
\(^2\)see e.g. Bernanke and Boivin (2003); Boivin and Ng (2005); D’Agostino and Giannone (2006); Forni, Hallin, Lippi, and Reichlin (2005, 2003); Giannone, Reichlin, and Sala (2004); Marcellino, Stock, and Watson (2003); Stock and Watson (2002a,b);
\(^3\)see e.g. Forni, Giannone, Lippi, and Reichlin (2005); Bernanke, Boivin, and Eliasz (2005); Giannone, Reichlin, and Sala (2004);
\(^4\)e.g. Chicago Fed National Activity Index or EuroCOIN, see Altissimo, Cristadoro, Forni, Lippi, and Veronese (2006) and Stock and Watson (1999), respectively;
and how it is related to their timeliness. The soft data include surveys and financial indicators and reflect market expectations. They are usually promptly available. In contrast, the hard indicators on real activity measure directly certain components of GDP (e.g. industrial production) and are published with a significant delay.

We propose several measures in order to assess the role of individual or groups of series in the forecast, while taking into account their respective publication lags. We find that surveys and financial data contain important information beyond the monthly real activity measures for the GDP forecasts, once their timeliness is properly accounted for.

The second chapter entitled *Maximum likelihood estimation of large factor model on datasets with arbitrary pattern of missing data* is based on a joint work with Michele Modugno. It proposes a methodology for the estimation of factor models on large cross-sections with a general pattern of missing data. In contrast to Giannone, Reichlin, and Small (2008), we can handle datasets that are not only characterised by a “ragged edge”, but can include e.g. mixed frequency or short history indicators. The latter is particularly relevant for the euro area or other young economies, for which many series have been compiled only since recently. We adopt the maximum likelihood approach, which, apart from the flexibility with regard to the pattern of missing data, is also more efficient and allows imposing restrictions on the parameters. Applied for small factor models by e.g. Geweke (1977), Sargent and Sims (1977) or Watson and Engle (1983), it has been shown by Doz, Giannone, and Reichlin (2006) to be consistent, robust and computationally feasible also in the case of large cross-sections. To circumvent the computational complexity of a direct likelihood maximisation in the case of large cross-section, Doz, Giannone, and Reichlin (2006) propose to use the iterative Expectation-Maximisation (EM) algorithm (used for the small model by Watson and Engle, 1983). Our contribution is to modify the EM steps to the case of missing data and to show how to augment the model in order to account for the serial correlation of the idiosyncratic component. In addition, we derive the link between the unexpected part of a data release and the forecast revision and illustrate how this can be used to understand the sources of the latter in the case of simultaneous releases.

We use this methodology for short-term forecasting and backdating of the euro area GDP on the basis of a large panel of monthly and quarterly data. In particular, we are able to examine the effect of quarterly variables and short history monthly series like the Purchasing Managers’ surveys on the forecast.

The third chapter is entitled *Large Bayesian VARs* and is based on a joint work with Domenico Giannone and Lucrezia Reichlin. It proposes an alternative approach to factor
models for dealing with the curse of dimensionality, namely Bayesian shrinkage.\footnote{In contrast to the previous chapters the real-time data flow is no longer considered.} We study Vector Autoregressions (VARs) which have the advantage over factor models in that they allow structural analysis in a natural way. We consider systems including more than 100 variables. This is the first application in the literature to estimate a VAR of this size. Apart from the forecast considerations, as argued above, the size of the information set can be also relevant for the structural analysis, see e.g. Bernanke, Boivin, and Eliasz (2005), Giannone and Reichlin (2006) or Christiano, Eichenbaum, and Evans (1999) for a discussion. In addition, many problems may require the study of the dynamics of many variables: many countries, sectors or regions. While we use standard priors as proposed by Litterman (1986), an important novelty of the work is that we set the overall tightness of the prior in relation to the model size. In this we follow the recommendation by De Mol, Giannone, and Reichlin (2008) who study the case of Bayesian regressions. They show that with increasing size of the model, one should shrink more to avoid overfitting, but when data are collinear one is still able to extract the relevant sample information. We apply this principle in the case of VARs.

We compare the large model with smaller systems in terms of forecasting performance and structural analysis of the effect of monetary policy shock. The results show that a standard Bayesian VAR model is an appropriate tool for large panels of data, once the degree of shrinkage is set in relation to the model size.

The fourth chapter entitled *Forecasting euro area inflation with wavelets: extracting information from real activity and money at different scales* proposes a framework for exploiting relationships between variables at different frequency bands in the context of forecasting. This work is motivated by the on-going debate whether money provides a reliable signal for the future price developments, see e.g. ECB (2008). The empirical evidence on the leading role of money for inflation in an out-of-sample forecast framework is not very strong, see e.g. Lenza (2006); Fischer, Lenza, Pill, and Reichlin (2008). At the same time, e.g. Gerlach (2003) or Assenmacher-Wesche and Gerlach (2007, 2008) argue that money and output could affect prices at different frequencies, however their analysis is performed in-sample. In this Chapter, it is investigated empirically which frequency bands and for which variables are the most relevant for the out-of-sample forecast of inflation when the information from prices, money and real activity is considered. To extract different frequency components from a series a wavelet transform is applied. It provides a simple and intuitive framework for band-pass filtering and allows a decomposition of series into different frequency bands. Its application in the multivariate out-of-sample forecast is novel in the literature. The results indicate that, indeed, different scales of money, prices and GDP can be relevant for the inflation...
References


