Journal of Banking & Finance 36 (2012) 2026–2047

Contents lists available at SciVerse ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Forecasting government bond yields with large Bayesian vector autoregressions

Andrea Carriero [a,*], George Kapetanios [a], Massimiliano Marcellino [b]

[a] School of Economics and Finance, Queen Mary University of London, Mile End Road, London E1 4NS, UK
[b] European University Institute, Bocconi University and CEPR, Villa San Paolo, via della Piazzuola 43, I-50133 Florence, Italy

ARTICLE INFO

Article history:
Received 9 June 2011
Accepted 7 March 2012
Available online 21 March 2012

JEL classification:
C11
C53
E43
E47

Keywords:
Bayesian methods
Forecasting
Marginal likelihood
Term structure

ABSTRACT

We propose a new approach to forecasting the term structure of interest rates, which allows to efficiently extract the information contained in a large panel of yields. In particular, we use a large Bayesian Vector Autoregression (BVAR) with an optimal amount of shrinkage towards univariate AR models. The optimal shrinkage is chosen by maximizing the Marginal Likelihood of the model. Focusing on the US, we provide an extensive study on the forecasting performance of the proposed model relative to most of the existing alternative specifications. While most of the existing evidence focuses on statistical measures of forecast accuracy, we also consider alternative measures based on trading schemes and portfolio allocation. We extensively check the robustness of our results, using different datasets and Monte Carlo simulations. We find that the proposed BVAR approach produces competitive forecasts, systematically more accurate than random walk forecasts, even though the gains are small.

1. Introduction

Producing accurate forecasts of the term structure of interest rates is crucial for bond portfolio management, derivatives pricing, and risk management. Unfortunately, all the forecasting models proposed so far in the macroeconomic and financial literature have a hard time in producing forecasts more accurate than a simple no-change forecast (i.e. a random walk forecast). The existing methods can be roughly categorized in three groups. The first two groups of models have the clear advantage of being grounded on finance theory, while the third group is the one that so far has produced the best results in out of sample forecast accuracy.

The first group contains models based on forward rate regressions. Such models try to forecast the future yields by extracting the information contained in the present forward rates. Prominent examples of this approach are, e.g. Fama and Bliss (1987) and Cochrane and Piazzesi (2005). Even though these papers document the existence of a predictive content in the forward rates, the out of sample forecasts produced by these models are typically outperformed by a simple no-change forecast (see e.g. Diebold and Li, 2006).

The second group contains models based on the No-Arbitrage paradigm. Typically the practical implementation of these models involves imposing an affine specification on a set of latent factors. Affine term structure models perform extremely well in fitting the yield curve in sample (see e.g. De Jong, 2000 and Dai and Singleton, 2000) but the performance in out of sample forecasting is quite poor. Duffee (2002) has shown that beating a random walk with a traditional no arbitrage affine term structure model is difficult. Ang and Piazzesi (2003) show that imposing no-arbitrage restrictions and an essentially affine specification of market prices of risk improves out-of-sample forecasts from a VAR(12), but the gain with respect to a random walk forecast is small. More favorable evidence in this respect has been found by Almeida and Vicente (2008). In this case one of the reasons for the difference in their results with respect to the rest of the literature (Duffee (2011), Joslin et al. (2011)) is that they also consider models with stochastic volatility while most of the literature only adopts Gaussian models. Also Favero et al. (2012) and Moench (2008) document a rather good performance of the ATSM models, but in both cases such models are complemented with a large macroeconomic dataset.

A third group of papers uses the Nelson and Siegel (1987) exponential components framework (Diebold and Li, 2006), possibly also imposing on it the no-arbitrage restrictions (Christensen et al., 2011). The forecasting results obtained by these models are better, with the Diebold and Li (2006) model producing one-year-ahead forecasts that are noticeably more accurate than standard benchmarks. Still, the gains are small at shorter forecast horizons.

In this paper we propose a new strategy for forecasting the term structure of interest rates, which exploits the information

* Corresponding author. Tel.: +44 20 7882 8050; fax: +44 20 8983 3580.
E-mail addresses: a.carriero@qmul.ac.uk (A. Carriero), g.kapetanios@qmul.ac.uk (G. Kapetanios), m.marcellino@eui.eu (M. Marcellino).

0378-4266/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jbankfin.2012.03.008

contained in a large panel of government bond yields. Focusing on the US, we show that our proposed strategy produces systematically better forecasts than all the methods outlined above. The method also outperforms the random walk, with small but systematic gains. The starting point of our strategy is the consideration that the yield curve can be thought of as a vector process composed of yields of different maturities. In that light, a straightforward approach to forecast would be to simply run a Vector Autoregression (VAR). However, such a strategy soon encounters the so-called "curse of dimensionality" problem, as the number of parameters to estimate rapidly reduces the degrees of freedom of the VAR system. As a result, the forecasts produced by a VAR are typically poor. To overcome this difficulty, one can either chose to sacrifice completely the cross-sectional information, and estimate e.g. a simple AR model on each yield, or try to summarize the information in an efficient but manageable manner. This latter possibility can be pursued by using a Bayesian Vector Autoregression (BVAR).

A BVAR is a VAR whose coefficients are random variables on which the researcher can impose some a priori information. Without entering into philosophical disputes about the Bayesian and the classical approach in econometrics, we think it is worth stressing here that BVARs can be interpreted simply as a selection device. Consider the first equation of a large VAR: there are many regressors, and the researcher needs to solve the trade-off between using as much information as possible, and the loss in degrees of freedom coming from having too many parameters to estimate. An intuitive way to proceed would be to start with an empty model, then adding a candidate regressor and performing a test of significance for that regressor. If the null is rejected, then the regressor is kept. The procedure can be repeated until the last candidate regressor is considered. What this procedure implicitly does is selecting the regressors on the basis of how much valuable information they contain. Information is valuable if it is able to significantly increase the likelihood of the model. The Bayesian algorithm works similarly. A priori the coefficient attached to a given candidate regressor is set to 0, and only if the information contained in the data is valuable enough to influence the likelihood the posterior mean will be far from 0. More precisely, rather than acting as a selection device which either includes or excludes a regressor, the BVAR includes all the regressors but it assigns a different weight to each of them. The weight is higher the higher is the informational content of a given regressor. The advantage of the BVAR over the simple step-wise procedure outlined above is that the former is a fully blown approach grounded on statistical theory, while the latter is not, and as a result it might lead to incorrect inference. For example, in the simple step-wise procedure the order with which the various candidate regressors are examined can significantly influence the final outcome, and the overall size of the step-wise procedure is unknown.

BVARs have a long story in econometrics. Although the good forecasting performance of BVARs has been documented years ago by Litterman (1986) and Doan et al. (1984), only recently they have started to be used more systematically for policy analysis and forecasting macroeconomic variables (Kadiyala and Karlsson, 1997; Banbura et al., 2010; Carriero et al., 2009, 2011a,b). One of the major stumbling blocks that prevented the use of BVARs as a model for forecasting and policy analysis has typically been the large computational burden they pose. Indeed, the computation of nonlinear functions of the parameters such as impulse-response functions and multi-step forecasts need to be performed via time consuming simulations. As we will discuss below in detail, in this paper we solve this problem by directly estimating the relevant coefficients for each forecast horizon, which allows us to compute the forecasts at all forecast horizons without resorting to simulation. As a result, the production of a full set of (Bayesian) forecasts for horizon 1- to 12-month ahead takes seconds.

We compare our proposed approach against all the major forecasting models used so far in the literature, including forward rate regression (Fama and Bliss (1987), Cochrane and Piazzesi (2005)), Affine term structure Models (Ang and Piazzesi, 2003), factor models (Stock and Watson, 2002a,b), and models based on the exponential components framework (Diebold and Li, 2006; Christensen et al., 2011). For reference, we include in the comparison also a set of linear models (random walk, autoregressive models, vector autoregressions). De Pooter et al. (2007) also propose to use Bayesian techniques to forecast the term structure, but their paper differs from ours as they use Bayesian model averaging over some term structure models, while we use a large BVAR to extract efficiently the information contained in a large cross section of yields.

Besides introducing the new approach to forecasting the yields, we extend the available empirical evidence in three directions: First, for all the models considered, we provide results using homogeneous datasets, while the existing results in the literature are based on various sample periods and selected maturities. We consider two alternative datasets of US yield curves: the Fama and Bliss (1987) unsmoothed yields, publicly available on the website of the Journal of Applied Econometrics, and the smoothed yields dataset by Gurkaynak et al. (2007), publicly available on the website http://www.federalreserve.gov. Secondly, while most of the existing evidence evaluates forecast accuracy only in terms of statistical measures such as Root Mean Squared Forecast Errors, we also evaluate forecasts on the basis of "economic" criteria. In particular, we provide Sharpe Ratios arising from simple trading rules based on the alternative forecasts, and we use the alternative forecasts to perform optimal portfolio allocation. Finally, we provide a simulation study in which we simulate and forecast a set of "artificial" term structures. In doing such a Monte Carlo simulation the researcher has to choose the Data Generating Process (DGP). Obviously the choice of a particular DGP over another would influence the results, advantaging one rather than the other model. Therefore, rather than concentrate on simulated data and an inevitably arbitrary data generation process, we carry out our simulation by bootstrapping the actual term structure dataset. The use of a real dataset as a basis for such a robustness analysis is referred to as a 'data based Monte Carlo method' and discussed further in, e.g. Ho and Sørensen (1996).

We find that: (i) our proposed BVAR approach produces forecasts systematically more accurate than the random walk forecasts; (ii) the gains with respect to the random walk are small; (iii) some models beat the BVAR for a few selected maturities and forecast horizons, but they perform much worse than the BVAR in the remaining cases; (iv) predictive gains with respect to the random walk have decreased over time; (v) different loss functions (i.e., "statistical" vs. "economic") lead to different ranking of specific models; (vi) modeling time variation in term premia is important and useful for forecasting.

The paper is structured as follows. Section 2 develops our BVAR approach. Section 3 introduces the competing forecasting models under comparison. Section 4 describes the data, the forecasting exercise and the alternative criteria we shall use in evaluating the alternative forecasts. Section 5 presents the main results, and Section 6 the robustness analysis. Finally, Section 7 concludes. To make the paper self-contained, Appendix A presents a set of technical derivations.

2. Bayesian VARs (BVAR)

The yield curve can be thought of as a vector process composed of yields of different maturities. In that light a straightforward approach to forecasting is to simply run a vector autoregression. However, such a strategy encounters an overparameterization problem, as the number of estimated parameters rapidly reduces the degrees of freedom of the VAR system. As a result, the forecasts