International Journal of Forecasting 30 (2014) 589–612

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Forecasting with factor-augmented error correction models

Anindya Banerjee[a,b,*], Massimiliano Marcellino[c,d,e], Igor Masten[f]

[a] Banque de France, France
[b] Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom
[c] European University Institute, Via della Piazzuola 43, 50133 Florence, Italy
[d] Bocconi University, Italy
[e] CEPR, United Kingdom
[f] University of Ljubljana, Faculty of Economics, Kardeljeva pl. 17, 1000 Ljubljana, Slovenia

ARTICLE INFO

Keywords:
Forecasting
Dynamic factor models
Error correction models
Cointegration
Factor-augmented error correction models
FAVAR

ABSTRACT

As a generalization of the factor-augmented VAR (FAVAR) and of the Error Correction Model (ECM), Banerjee and Marcellino (2009) introduced the Factor-augmented Error Correction Model (FECM). The FECM combines error-correction, cointegration and dynamic factor models, and has several conceptual advantages over the standard ECM and FAVAR models. In particular, it uses a larger dataset than the ECM and incorporates the long-run information which the FAVAR is missing because of its specification in differences. In this paper, we examine the forecasting performance of the FECM by means of an analytical example, Monte Carlo simulations and several empirical applications. We show that FECM generally offers a higher forecasting precision relative to the FAVAR, and marks a useful step forward for forecasting with large datasets.

1. Introduction

Banerjee and Marcellino (2009) introduced the Factor-augmented Error Correction Model (FECM). The paper's main contribution was to bring together two important recent strands of the econometric literature on modelling co-movements, which had a common origin but had thus far remained largely separate in their implementation, namely, cointegration and dynamic factor models. The focus was on a theoretical framework that allowed for the explicit introduction of cointegrating or long-run information into a dynamic factor model, and evaluated the influence of incorporating long-run information in modelling data, particularly in situations where the dataset available to researchers was potentially very large (as in the empirical illustrations described in Section 5 below). It was argued that the FECM, where the factors extracted from the large dataset are modelled jointly with a limited set of economic variables of interest, represented a manageable way of dealing with the problem posed by large datasets characterized by cointegration, where, in principle, such cointegration needs to be taken into account. A number of papers, see for example that by Clements and Hendry (1995), have emphasized the complexity of modelling large systems of equations where the complete cointegrating space may be difficult to identify. Therefore, proxying for the missing cointegrating information by using factors could turn out to be extremely useful, and the FECM was proposed as a potentially worthwhile approach, with applicability to a wide range of situations.

The discussion by Banerjee and Marcellino (2009) concentrated on first establishing a theoretical structure for describing the FECM, then illustrating its efficacy by the use of analytical examples, a simulation study and

* Corresponding author at: Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. Tel.: +44 778 098 0293.
E-mail addresses: a.banerjee@bham.ac.uk (A. Banerjee), massimiliano.marcellino@eui.eu (M. Marcellino), igor.masten@ef.uni-lj.si (I. Masten).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.01.009

two empirical applications. The model comparisons were based mainly on in-sample measures of model fit, and the improvements provided by FECMs were studied with respect to a standard Error Correction Model (ECM) and Factor-Augmented VARs (FAVAR) such as those considered by Bernanke, Boivin, and Eliasz (2005), Favero, Marcellino, and Neglia (2005) and Stock and Watson (2005). The FECM was viewed as an improvement over both the ECM, by relaxing the dependence of cointegration analysis on a small set of variables, and the FAVAR, by allowing for the inclusion of error correction terms in the equations for the key variables under analysis, thus preventing the errors from being non-invertible MA processes.

The focus of this paper is instead upon the evaluation of the forecasting performance of the FECM in comparison with those of the ECM and the FAVAR. In our view, establishing forecasting efficacy is an important key to determining the usefulness of the FECM as an econometric tool. As we show below, the relative rankings of the ECM, the FECM and the FAVAR depend upon the variables being modelled and the features of the processes generating the data, such as the amount and strength of any cointegration, the degree of lagged dependence in the models, and the forecasting horizon. However, in general, the FECM tends to perform better than either the ECM or the FAVAR, given that the former is a nesting specification.

We start in Section 2 by reviewing the theoretical background of our study, including describing the FECM and comparing it with the ECM and the FAVAR.

Section 3 offers a simple yet comprehensive analytical example to assist our understanding of the features which are likely to determine the rankings — in terms of forecasting accuracy — of these three models.

Section 4 presents two Monte Carlo designs for illustrating the effectiveness of the different models in providing forecasts. The first design is based on the simple analytical model of Section 3, while the second is more elaborate and mimics one of the models estimated in the empirical examples given in Section 5. We expect that the results of the Monte Carlo will show that the strength of the error correction, along with the lengths of the cross-section (N) and the time dimension (T), are highly important in determining the forecast rankings of alternative models. However, the FECM performs well in the majority of cases, and is generally better than the FAVAR.

Section 5 carries the analysis to the practical realm. Forecasting with ECMs and factor models has attracted a considerable amount of attention, see for example Clements and Hendry (1995) and Eickmeier and Ziegler (2008), respectively. To provide a thorough comparison of the ECM, FAVAR and FECM, we consider six main applications, which are described briefly in turn below.

Our first two applications are related to the work of Stock and Watson (2002b), who focused on forecasting (a) a set of four real variables (total industrial production, personal income less transfers, employment on non-agricultural payrolls, and real manufacturing trade and sales — all of which are relevant to the assessment of business cycle conditions) and (b) a set of four nominal variables (inflation of producer prices of finished goods, CPI inflation with all items included, CPI inflation less food, and the growth of the personal consumption expenditure deflator) for the United States (US). They compared the performances of factor models, ARs and VARs, and typically found gains from the use of factor models. Since the four variables in each set represent strongly related economic phenomena, it is logical to expect that they will be cointegrated. Hence, in this context, the FECM represents a natural econometric specification and one whose usefulness we investigate in Section 5 below.

Our third and fourth applications focus on small monetary systems, consisting of one real, one nominal and one financial variable, as is standard practice in this literature, see e.g. Rudebusch and Svensson (1998). Favero et al. (2005), among others, considered augmenting this model with factors extracted from a large dataset, to assess the effects on estimation and shock transmission. Here, we are more interested in forecasting, and in the role of cointegration both among these basic variables, and between the basic variables and the factors. The VAR, FECM and FAVAR models are estimated first for the US using the same dataset as Stock and Watson above, then for Germany, the largest country in the euro area, for which much shorter time series are available, due to unification.

The fifth application concerns the term structure of interest rates. A standard model for these variables assumes that they are driven by three factors, namely the intercept, slope and curvature, see e.g. Diebold and Li (2006). Hence, there should be a large amount of cointegration among them, in line with the findings of Hall, Anderson, and Granger (1992). Therefore, the FECM should be well suited to this context.

The sixth and final application deals with exchange rate forecasting. Both the empirical analysis of Meese and Rogoff (1983) and the theoretical results of Engel and West (2005), among others in this vast literature, point to the difficulties of beating a random walk or simple AR forecast. However, Carriero, Kapetanios, and Marcellino (2009) show that cross-sectional information can be useful, but that factor models on their own do not appear to work very well in forecasting. Since this poor performance could be due to the omission of information relating to cointegration, in this sense, FECMs are another obvious set of candidates to try in this framework.

Due to space constraints, only the first two empirical applications have detailed results presented in this paper, with the results for the remaining four being summarized in Section 5.2. However, the complete sets of results are available in the Appendix on-line, together with some further Monte Carlo results and full details of the datasets used.[1]

Before proceeding further, it is helpful to begin by highlighting the key results of this extensive empirical analysis. First, for real (business cycle related) variables for the US, the FECM tends to perform better than either the FAVAR or the ECM. Second, for the nominal US variables, either an adaptation of FECM which includes stationary factors, denoted FECMc (discussed below), or the ECM are generally the preferred models (depending upon the time

[1] These are of course also available from us upon request.