International Journal of Forecasting 17 (2001) 585–605

international journal of forecasting

www.elsevier.com/locate/ijforecast

Bayesian prediction with cointegrated vector autoregressions

Mattias Villani

Department of Statistics, Stockholm University, S-106 91 Stockholm, Sweden

Abstract

A complete procedure for calculating the joint predictive distribution of future observations based on the cointegrated vector autoregression is presented. The large degree of uncertainty in the choice of cointegration vectors is incorporated into the analysis via the prior distribution. This prior has the effect of weighing the predictive distributions based on the models with different cointegration vectors into an overall predictive distribution. The ideas of Litterman [Mimeo, Massachusetts Institute of Technology, 1980] are adopted for the prior on the short run dynamics of the process resulting in a prior which only depends on a few hyperparameters. A straightforward numerical evaluation of the predictive distribution based on Gibbs sampling is proposed. The prediction procedure is applied to a seven-variable system with a focus on forecasting Swedish inflation.

Keywords: Bayesian; Cointegration; Inflation forecasting; Model averaging; Predictive density

1. Introduction

The idea of cointegration (Engle & Granger, 1987; Stock & Watson, 1988) has become extremely popular in applied work. With the introduction of this concept, econometricians were given the possibility to incorporate theoretically motivated long run equilibriums into their otherwise relatively unrestricted models. It was expected that such long run restrictions would lead to improved forecasting[1] ability, especially for longer forecasting horizons (Engle & Yoo, 1987), and such findings have indeed been made; see LeSage (1990), Shoesmith (1992, 1995) and Stark (1998).

In most applications, many theoretically motivated cointegration restrictions are available and the choice between them is often difficult. Even though the restrictions are empirically testable, the evidence from such tests may be inconclusive; we may obtain weak support for one or several of the theoretical restrictions while other restrictions are clearly rejected or we may even find strong support for more than one of them. In addition, we must also consider the possibility to ignore them all and estimate the long run relations entirely from data.

If the model is to be used for forecasting then

*Tel.: +46-8-162578; fax: +46-8-167511.

E-mail address: mattias.villani@stat.su.se (M. Villani).

[1]The two terms forecasting and prediction will be treated as synonyms even though the former is usually restricted to future events, see Clements and Hendry (1998, p. 39).

0169-2070/01/$ – see front matter
PII: S0169-2070(01)00082-6

one can produce separate forecasts from each of the models with the different theoretical restrictions imposed and from the model with empirically estimated restrictions. It seems reasonable, however, to weight these forecasts by weights proportional to the empirical support of the restriction. In any case, it is often unsatisfying, and practically complicated, to have a whole range of forecasts based on different restrictions.

The aim of this paper is to introduce a complete Bayesian approach to prediction using a cointegrated VAR model. In a Bayesian setting, the uncertainty regarding the correct cointegration restriction should be reflected in the prior distribution of the model parameters. We propose a prior which allows us to work simultaneously with several plausible cointegration restrictions and provides an optimal way to combine the forecasts. Furthermore, the suggested prior also allows some or even all of the cointegration restrictions to be estimated freely from data. In effect, theoretically motivated long run restrictions are imposed only if they are supported by data.

Even if we impose long-run restrictions on the VAR, it is usually the case that most parameters describe the short run dynamics of the system and these parameters are still unrestricted. As VAR models usually have many parameters, unrestricted estimation of the model parameters can be very imprecise. Instead of restricting some of the parameters in the usual way, Litterman (1980, 1986) introduced a prior distribution expressing the belief (soft restrictions) that the variables in the system are most likely to be independent random walks. A slight variant of this idea is used here.

An often neglected part of the forecasting phase is the uncertainty in the point forecasts, which in many cases is as important as the point forecasts. In traditional forecasting applications of the VAR with cointegration restrictions, the forecasting uncertainty is distorted primarily by two conditionings. Firstly, to impose restrictions on the cointegration relations when it is uncertain whether or not such restrictions hold is to ignore an important part of the model uncertainty; this will result in misleading reports of forecasting uncertainty. Secondly, the uncertainty attributed to the estimation of the model parameters is at best accounted for by rough, and often dubious, corrections based on asymptotic theory. The aim here is to produce uncertainty statements that fully account for all sources of uncertainty and not only those attributed to future stochastic disturbances to the system.

2. The cointegrated vector autoregression

Consider the ordinary p-dimensional vector autoregressive process with K lags

$$\mathbf{x}_t = \sum_{i=1}^{K} \mathbf{\Pi}_i \mathbf{x}_{t-i} + \mathbf{\Phi d}_t + \boldsymbol{\varepsilon}_t, \qquad (2.1)$$

where $\mathbf{x}_t$ $(p \times 1)$ contains an observation on the p time series at time t, $\mathbf{\Pi}_i$ $(p \times p)$ is the matrix of coefficients describing the dynamics of the system while $\mathbf{d}_t$ $(d \times 1)$ contains d deterministic trend or seasonal dummy variables at time t whose effect on $\mathbf{x}_t$ is captured by $\mathbf{\Phi}$ $(p \times d)$. Finally, $\boldsymbol{\varepsilon}_t$ $(p \times 1)$ is a vector of error terms assumed to be distributed as $N_p(\mathbf{0}, \mathbf{\Sigma})$ distribution with independence between time periods.

An isomorphic parametrization of the VAR model, better suited for cointegration analysis, is the error correction model (ECM)

$$\Delta\mathbf{x}_t = \mathbf{\Pi x}_{t-1} + \sum_{i=1}^{K-1} \mathbf{\Gamma}_i \Delta\mathbf{x}_{t-i} + \mathbf{\Phi d}_t + \boldsymbol{\varepsilon}_t, \qquad (2.2)$$

where $\Delta\mathbf{x}_{t-i} = \mathbf{x}_{t-i} - \mathbf{x}_{t-i-1}$, $\mathbf{\Pi} = \Sigma_{i=1}^{K} \mathbf{\Pi}_i - \mathbf{I}_p$ and $\mathbf{\Gamma}_i = -\Sigma_{j=i+1}^{K} \mathbf{\Pi}_j$. The greatest merit of the ECM is that it separates the long-run component of the series $(\mathbf{\Pi x}_{t-1})$ from the short-run dynamics $(\Sigma_{i=1}^{K-1} \mathbf{\Gamma}_i \Delta\mathbf{x}_{t-i})$, a separation that both simplifies the interpretation of the non-station-