Journal of Econometrics 178 (2014) 779–793

Contents lists available at ScienceDirect

Journal of Econometrics

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

Time-varying sparsity in dynamic regression models

Maria Kalli [a,1], Jim E. Griffin [b,*]

[a] Canterbury Christ Church University Business School, Canterbury Christ Church University, Canterbury, UK

[b] School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK

ARTICLE INFO

Article history:
Received 21 February 2012
Received in revised form
23 August 2013
Accepted 31 October 2013
Available online 8 November 2013

JEL classification:
C11
C15
C52
C53
C58

Keywords:
Time-varying regression
Shrinkage priors
Normal-gamma priors
Markov chain Monte Carlo
Equity premium
Inflation

ABSTRACT

A novel Bayesian method for inference in dynamic regression models is proposed where both the values of the regression coefficients and the importance of the variables are allowed to change over time. We focus on forecasting and so the parsimony of the model is important for good performance. A prior is developed which allows the shrinkage of the regression coefficients to suitably change over time and an efficient Markov chain Monte Carlo method for posterior inference is described. The new method is applied to two forecasting problems in econometrics: equity premium prediction and inflation forecasting. The results show that this method outperforms current competing Bayesian methods.

1. Introduction

Forecasting, the estimation of a future value of a variable, plays an important role in both decision-making and strategic planning and has been extensively studied in econometrics. For example, forecasts of inflation affect the decisions of monetary and fiscal policymakers, investors who wish to hedge against the risk of nominal assets, and trade unions and management when they negotiate wage contracts, to name a few. Similarly, forecasts of equity premiums play an important role for investors who wish to diversify their equity portfolios to hedge against adverse market movements. The quality of the forecast depends on: the time scale involved (how far into the future we are trying to predict), the time period of the empirical sample, and the model used.

Regression models are a popular technique for forecasting since the values of other variables can be used to inform predictions. However, their use with observations made over time is complicated by several problems. Firstly, it has been found that these models can produce poor out-of-sample forecasts when the predictors' effects are assumed constant over time. This is generally taken as evidence that the effects of variables are time-varying. Sims (1980), Stock and Watson (1996), Cogley and Sargent (2001, 2005), Primicery (2005), Paye and Timmermann (2006), Ang and Bekaert (2007), Canova (2007), and Lettau and Van Nieuwerburgh (2008) are some studies providing evidence of time varying regressor effects in inflation and equity premium forecasting. Secondly, the increasing availability of large economic datasets has led to interest in using regression models with many regressors. It is well known that the estimation of regression models becomes more complicated when a large number of predictors is used due to the increased potential for over-fitting which can lead to poor out-of-sample forecasts or predictions. The problem of over-fitting can be alleviated by looking for sparse regression estimates where many regression coefficients are set to zero or values close to zero. This is usually achieved using regularisation of the regression coefficients or variable selection methods.

The problem of time-varying regressor effects can be addressed using dynamic regression models, which are a form of time-varying parameter models, where the regression coefficients are

* Corresponding author. Tel.: +44 1227 82 3865; fax: +44 1227 82 7932.
E-mail addresses: maria.kalli@canterbury.ac.uk (M. Kalli), J.E.Griffin-28@kent.ac.uk (J.E. Griffin).

1 Tel.: +44 1227863027; fax: +44 1227827932.

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.10.012

assumed to evolve according to some stochastic process. This defines the dynamic linear models (DLM) discussed in West and Harrison (1999), or state-space models.

The problem of a large number of variables has been addressed in several ways. Initial work concentrated on models which assume a global measure of the importance of a variable. Groen et al. (2009) introduced a latent variable which indicates whether a variable is included in or excluded from the model. The approach is restricted so that the decision to include or exclude a predictor is irreversible. Belmonte et al. (2011) combined the Bayesian Lasso of Park and Casella (2008) with the model selection methods of Frühwirth-Schnatter and Wagner (2010) in order to have shrinkage in a dynamic regression setting. This approach allows some regression coefficients to be shrunk very close to zero for the whole time series and so effectively achieve variable selection. These methods have the potentially critical limitation that the importance of variables cannot change over time. For example, in some problems certain predictors could be useful for forecasting at particular times but not at others. In the Bayesian literature, this problem has been approached by allowing variables to enter and exit the model over time. Koop and Korobilis (2012) used the dynamic model averaging (DMA) method of Raftery et al. (2010) to select a suitable time-varying parameter model. However, the dynamics on model space are only implicitly defined by their approach. Alternatively, Chan et al. (2012) constructed the class of time-varying dimension (TVD) models using an explicitly constructed stochastic process for the subset of variables included in the model. This leads to a dynamic mixture model for which efficient posterior computational methods can be developed using the approach of Gerlach et al. (2000). However, both approaches are limited by the number of models that they can consider. The DMA approach uses full enumeration of posterior probabilities and the number of models with p regressors is 2^p precluding large values of p. Posterior computation in the TVD model also potentially involves all 2^p models but the authors suggest using a much restricted set of possible models.

The DMA and TVD approaches build on Bayesian variable selection techniques which explicitly consider all possible regression models. An alternative class of methods is Bayesian regularisation methods which use absolutely continuous priors and encourage small regression effects to be aggressively shrunk towards zero under the posterior (see Carvalho et al. (2010), and Polson and Scott (2011)). These authors have shown that these methods can lead to posteriors which place substantial mass on combinations of regression coefficients which are sparse (that is most of the regression coefficients have values very close to zero) if supported by the data. Belmonte et al. (2011) have already extended one such prior, the Bayesian Lasso, to the dynamic regression setting. Our methodological contribution differs from their work in two main respects. Firstly, our prior for the time-varying regression coefficients extends the more general normal-gamma (NG) prior (see Caron and Doucet (2008) and Griffin and Brown (2010)) to DR models and, secondly, our prior accounts for both time-varying regression coefficients and time-varying sparsity.

The paper is organised as follows: Section 2 introduces the normal-gamma autoregressive (NGAR) process prior, considers some of its properties and describes the full Bayesian model for dynamic regression with time-varying sparsity. Section 3 provides of an overview of the required Markov chain Monte Carlo (MCMC) method for fitting a dynamic regression model with an NGAR process prior (the full steps of the MCMC sampler are described in Appendix A). Section 4 applies the DR model with NGAR process priors to simulated data, while Section 5 considers empirical studies in equity premium prediction and inflation forecasting. Section 6 summarises our findings and conclusions.

2. A Bayesian dynamic regression model with time-varying sparsity

A dynamic regression (DR) model links a response y_t to regressors $x_{1,t}, \ldots, x_{m,t}$ (all observed at time t) by

$$y_t = \sum_{i=0}^{m} x_{i,t}\beta_{i,t} + \epsilon_t, \quad t = 1, \ldots, T,\ i = 0, \ldots, m \tag{1}$$

where $x_{0,t} = 1$ for all t (allowing for an intercept), $\beta_{i,t}$ is a vector of unknown coefficients for the ith regressor at time t, ϵ_t is the innovation term at time t generated from a normal distribution with zero mean and time-varying variance i.e. $\epsilon_t \sim \mathrm{N}(0, \sigma_t^2)$. The regressors $x_{1,t}, \ldots, x_{m,t}$ may include both lags of the response and exogenous variables. The DR model is usually completed by assuming that $\beta_{1,t}, \ldots, \beta_{m,t}$ follow a linear stochastic process (such as a random walk or vector autoregression).

In this paper we assume that the time-varying variances, $\sigma_1^2, \ldots, \sigma_T^2$ are generated by a gamma autoregressive (GAR) process using the method described in Pitt et al. (2002) and Pitt and Walker (2005) and later, independently, developed as the autoregressive gamma process by Gourieroux and Jasiak (2006). The process is specified using latent variables $\kappa_1^\sigma, \ldots, \kappa_{T-1}^\sigma$ by the recursion

$$\sigma_t^2 \sim \mathrm{Ga}\left(\lambda^\sigma + \kappa_{t-1}^\sigma, \lambda^\sigma / \left(\mu^\sigma \left(1 - \rho^\sigma\right)\right)\right) \quad \text{and}$$

$$\kappa_{t-1}^\sigma | \sigma_{t-1}^2 \sim \mathrm{Pn}\left(\lambda^\sigma \rho^\sigma \sigma_{t-1}^2 / \left(\left(1 - \rho^\sigma\right) \mu^\sigma\right)\right),$$

for $t = 2, \ldots, T$ and $\sigma_1^2 \sim \mathrm{Ga}\left(\lambda^\sigma, \lambda^\sigma/\mu^\sigma\right)$. This defines a first-order autoregressive model for $\sigma_1^2, \ldots, \sigma_T^2$ with autoregressive parameter ρ^σ and stationary distribution $\mathrm{Ga}(\lambda^\sigma, \lambda^\sigma/\mu^\sigma)$ where $x \sim \mathrm{Ga}(a, b)$ denotes that x follows a gamma distribution with shape parameter a and mean a/b. We discuss our choice of priors for $\lambda^\sigma, \mu^\sigma,$ and ρ^σ in Section 2.2. The Bayesian model is completed by specifying a prior for $\beta_{i,t}$ for $i = 0, \ldots, m$ and $t = 1, \ldots, T$ which is discussed in the following section.

2.1. The NGAR process prior for $\beta_{i,t}$

In regression models with a large number of regressors, it is common to assume that only a subset of the regressors is important for prediction. In DR models, this assumption is naturally extended to subsets of important regressors that change over time. This assumption can be expressed in the prior by defining a stochastic process for $\beta_{1,t}, \ldots, \beta_{m,t}$ which allows a subset of $\beta_{1,t}, \ldots, \beta_{m,t}$ to be set equal to zero (or equivalently, some regressors to be removed from the model), or values close to zero at time t and allows the subset to change over time. We refer to the proportion of parameters $\delta = (\delta_1, \ldots, \delta_s)$ which are close to zero as the sparsity of δ with a larger proportion referred to as more sparsity. In DR models, there are two interesting forms of sparsity. Firstly, the sparsity of $\beta_i = (\beta_{i,1}, \ldots, \beta_{i,T})$ which is the proportion of time that $\beta_{i,t}$ is close to zero. Secondly, the sparsity of $\beta_{1,t}, \ldots, \beta_{m,t}$ which is the proportion of regression coefficients that are set close to zero at time t. The assumption of time-varying subsets of important variables can be expressed by time-varying sparsity of $\beta_{1,t}, \ldots, \beta_{m,t}$.

These forms of sparsity can be expressed by giving independent normal-gamma autoregressive (NGAR) process priors to the time series of regression coefficients $\beta_1, \ldots, \beta_m$. We define the NGAR process prior below. Let $x \sim \mathrm{Pn}(\mu)$ denote that x follows a Poisson distribution with mean μ.

Definition 1. The normal-gamma autoregressive (NGAR) process for β_i is defined by

$$\beta_{i,s} = \sqrt{\frac{\psi_{i,s}}{\psi_{i,s-1}}}\varphi_i \beta_{i,s-1} + \eta_{i,s},$$

$$\eta_{i,s}|\psi_{i,s} \sim \mathrm{N}\left(0, (1 - \varphi_i^2)\psi_{i,s}\right) \quad s = 2, \ldots, T,$$