Time-Varying Parameter Models – Achieving Shrinkage and Variable Selection

Angela Bitto[1], Sylvia Frühwirth-Schnatter[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 4

[1] WU Vienna University of Economics and Business, Institute for Statistics and Mathematics, Vienna, Austria
`angela.bitto@wu.ac.at`,
`sylvia.fruehwirth-schnatter@wu.ac.at`

Abstract

The present paper contributes to the literature in two ways. First, we investigate shrinkage for Time-Varying Parameter (TVP) models based on the Normal-Gamma prior which has been introduced by [2] for standard regression models. Our approach extends [1] who considered the Bayesian LASSO prior, a special case of the Normal Gamma prior. While both priors reduce the risk of overfitting and increase statistical efficiency, they do not allow for variable selection. Hence, as a second contribution, we follow [3] and consider TVP models with spike-and-slab priors which explicitly incorporate variable selection both with respect to the initial parameters as well as their variances. Following [1] we choose EU area inflation modelling based on the generalized Phillips curve as our application.

Keywords: Time-varying parameter model; Hierarchical Shrinkage; Spike-and-slab; Normal-Gamma prior.

1 Introduction

Time-varying parameter (TVP) models are a popular tool for handling data with smoothly changing parameters, with initial value $(\beta_1, \ldots, \beta_d)$:

$$y_t = x_t\boldsymbol{\beta}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}\left(0, \sigma_t^2\right), \tag{1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Q}\right), \tag{2}$$

$$\boldsymbol{Q} = Diag(\theta_1, \ldots, \theta_d). \tag{3}$$

However, in situations with many parameters the flexibility underlying these models may lead to overfitting models and, as a consequence, to a severe loss of statistical efficiency. This occurs, in particular, if only a few parameters are truely time-varying, while the remaining ones are constant or even insignificant. As a remedy, hierarchical shrinkage priors have been introduced for TVP models to allow shrinkage both of the initial parameters β_j as well as their variances θ_j towards zero.

2 Hierarchical Shrinkage Priors

We consider the Normal-Gamma prior which has been introduced by [2] as a shrinkage prior, both for the initial values and the square root of the variances:

$$\beta_j \sim \mathcal{N}\left(0, \tau_j^2\right), \qquad \tau_j^2 \sim \mathcal{G}\left(a^\tau, a^\tau \lambda^2/2\right), \qquad \lambda^2 \sim \mathcal{G}\left(d_1, d_2\right) \tag{4}$$

$$\sqrt{\theta}_j \sim \mathcal{N}\left(0, \xi_j^2\right), \qquad \xi_j^2 \sim \mathcal{G}\left(a^\xi, a^\xi \kappa^2/2\right), \qquad \kappa^2 \sim \mathcal{G}\left(e_1, e_2\right) \tag{5}$$

For $a^\tau = 1$ and $a^\xi = 1$, the Lapalce prior results. Crucial for both shrinkage priors is the choice of the hyperparameters, in particular the choice of d_1, d_2, e_1, e_2. We dedicate an extensive analysis to this topic.

3 Spike-and-slab Priors

The ability to sort each regressor into one of the following categories - either *(1)dynamic, (2)significant, but static or (3) not significant at all* - is very often of interest. As none of the priors discussed above can easily establish this classification, we orient the second part of this paper to a methodoly which allows for variable selection. For a formal classification we follow [3] and use a spike-and-slab prior both for the initial values as well as for the variances.

A spike-and-slab prior is a finite mixture distribution with two components, where one component (the spike) has much stronger global shrinkage than the second component (the slab):

$$\beta_j \sim (1-\omega_\delta)\Delta_0(\beta_j) + \omega_\delta p_{slab}(\beta_j), \quad \beta_j \sim \mathcal{N}\left(0, 1/\beta^*\right) \tag{6}$$

$$\sqrt{\theta_j} \sim (1-\omega_\gamma)\Delta_0(\sqrt{\theta_j}) + \omega_\gamma p_{slab}(\sqrt{\theta_j}), \quad \sqrt{\theta}_j \sim \mathcal{N}\left(0, 1/\theta^*\right), \tag{7}$$

where Δ_0 is a Dirac spike at 0 and p_{slab} is a normal prior. The finite mixture structure allows to introduce, respectively, binary indicators δ_j and γ_j to classify each initial value β_j and each variance θ_j into one of the two components.

4 Application

Following [1], hierarchical shrinkage priors as well as spike-and-slab priors are applied to EU area inflation modelling based on the generalized Phillips curve.

In this context inflation depends on lags of inflation and other predictors such as unemployment rate, money supply and changes in the oil price. We make use of the same dataset consisting of EU monthly data from February 1994 until November 2010.

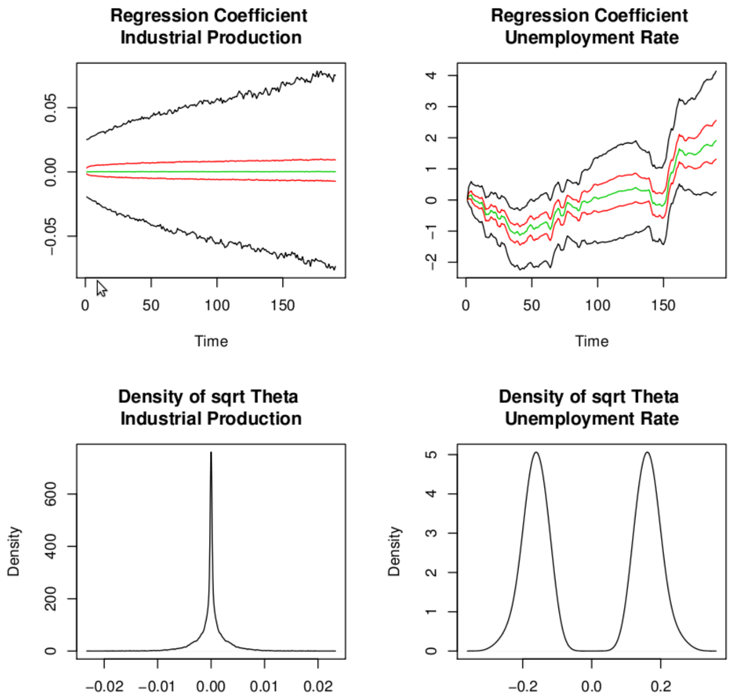

Figure 1: Paths and posterior densities for unemployment rate and industrial production

Figure 1 shows the posterior paths of β_{tj} as well as the posterior density of $\sqrt{\theta_j}$ for two predictors (unemployment rate and industrial production). An exploratory analysis suggests that industrial production is neither significant nor time-varying. However, for the unemployment rate it is less obvious whether this predictor is significant or time-varying.

In contrast, our second approach based on spike-and-slab priors allows for intrinsic classification. As shown in Table 1, we find for some predictors, that the regression coefficient is cleary time-varying (Intercept, Money Supply, Unemployment Rate), while other predictors are definitely insignificant (Economic Sentiment Indicator, Industrial Production) and this result is robust to the choice of the hyperparameters β^* and θ^*. However, as for hierarchical shrinkage priors, for some predictors, such as 1m Euribor and 1y Euribor, it is difficult to decide whether they are significant or time-varying, because this decision is strongly influenced by the choice of the hyperparameters.

Since the corresponding time series are relatively short, the various priors

Probability $Pr(\delta_i = 1 \vert data)$ & $Pr(\gamma_i = 1 \vert data)$						
Prior	$\beta^*, \theta^* = 0.1$		$\beta^* = 0.1$, $\theta^* = 1$		$\beta^* = 1$, $\theta^* = 10$	
	$\delta_i = 1$	$\gamma_i = 1$	$\delta_i = 1$	$\gamma_i = 1$	$\delta_i = 1$	$\gamma_i = 1$
Intercept	1.00	1.00	1.00	1.00	1.00	0.90
Money Supply	1.00	1.00	1.00	1.00	1.00	1.00
Unemployment Rate	1.00	1.00	1.00	1.00	1.00	1.00
Economic Sentiment Indicator	0.20	0.01	0.20	0.01	0.05	0.00
Industrial Production	0.15	0.00	0.24	0.00	0.04	0.00
1m Euribor	0.52	0.00	1.00	1.00	0.17	0.01
1y Euribor	0.94	0.51	0.67	0.02	0.17	0.01

Table 1: Classification results based on spike-and-slab priors for various hyperparameters β^* and θ^*; ω_γ and $\omega_\delta \sim$ Uniform; 40,000 draws, burn-in: 20,000.

lead to quite different conclusions. Inference considerably depends on the particular choice of hyperparameters.

References

[1] M. A. G. Belmonte, G. Koop and D. Korobilis. **Hierarchical shrinkage in time varying parameter models**. *Technical Report, University of Strathclyde, Glasgow*; 2011.

[2] J. E. Griffin and P. J. Brown. **Inference with normal-gamma prior distributions in regression problems**. *Bayesian Analysis 5*; 2010; pp. 151-170.

[3] S. Frühwirth-Schnatter and H. Wagner. **Stochastic model specification search for Gaussian and partially non-Gaussian state space models**. *Journal of Econometrics 154*; 2010; pp. 85-100.