Economics Letters 133 (2015) 35–39

Contents lists available at ScienceDirect

Economics Letters

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

Flexible model comparison of unobserved components models using particle Gibbs with ancestor sampling

Nima Nonejad *

University of Rome "Tor Vergata", Italy
CREATES, Denmark

HIGHLIGHTS

- Unobserved components models with stochastic volatility (SV) effects are widely used to model inflation rates.
- However, formal model comparison using Gibbs sampling is difficult.
- We show that PG-AS provides a flexible framework for estimation and model comparison.
- We provide applications using US and UK data, comparing different models.
- The model with time-varying SV in mean effects performs best.

ARTICLE INFO

Article history:
Received 5 March 2015
Received in revised form
27 April 2015
Accepted 29 April 2015
Available online 7 May 2015

JEL classification:
C11
C22
C52
C63

Keywords:
Gibbs sampling
Particle filtering
Model comparison

ABSTRACT

In this paper, we show that particle Gibbs with ancestor sampling (PG-AS) provides a unified and flexible framework for estimation and model comparison of unobserved components models with time-varying volatility effects, which are widely used in inflation rate modeling.

1. Introduction

Unobserved components (UC) models with time-varying volatility modeled as a stochastic volatility (SV) process are widely used in economics, especially in the context of inflation rate modeling. Over the years, many different varieties of the UC model with SV effects have been proposed, see Stock and Watson (2007), Grassi and Proietti (2010) and Chan (2013, 2014). Naturally, it is interesting to investigate which specification fits the data best. However, model comparison using marginal likelihood (ML) or deviance information (DIC) criteria is rarely done in practical applications. This is because obtaining these quantities is cumbersome, often requiring additional computational effort, see for instance Grassi and Proietti (2010).

In this paper, we show that particle Gibbs with ancestor sampling (PG-AS), suggested in Lindsten et al. (2014) provides a very flexible framework for estimation and especially ML and DIC computation of UC models with SV and SV in mean effects. We start by estimating an UC model where the stochastic volatility process has a direct and time-varying impact on the inflation rate. Furthermore, the integrated likelihood is directly available through the conditional particle filter with ancestor sampling (CPF-AS), making ML and DIC computation straightforward. Other varieties of the UC model, including a specification with SV effects in both the transitory and core components of inflation are also estimated using PG-AS. Systematic model selection using ML and DIC provides evidence in favor of the UC model with time-varying SV in mean (UC-SVM) effects.

* Correspondence to: Department of Economics and Finance, University of Rome "Tor Vergata", Via Columbia 2, 00133 Rome, Italy.
E-mail address: nimanonejad@gmail.com.

http://dx.doi.org/10.1016/j.econlet.2015.04.034
0165-1765/

The rest of this paper is organized as follows: Section 2 introduces the models. PG-AS estimation and results using quarterly US and UK inflation data are presented in Sections 3 and 4. The last section concludes. An Appendix at the end of the paper provides technical details on CPF-AS.

2. Models

Our initial model is the UC-SVM model, see also Chan (2014). This model is given as

$$y_t = \mu_t + \alpha_t \exp(h_t) + \varepsilon_t^y, \quad \varepsilon_t^y \sim N(0, \exp(h_t)) \tag{2.1}$$

$$h_t = \mu_h + \phi_h (h_{t-1} - \mu_h) + \varepsilon_t^h, \quad \varepsilon_t^h \sim N\left(0, \sigma_h^2\right) \tag{2.2}$$

$$\gamma_t = \gamma_{t-1} + \varepsilon_t^\gamma, \quad \varepsilon_t^\gamma \sim N(0, \Omega), \tag{2.3}$$

where $\gamma_t = (\mu_t, \alpha_t)'$, $\Omega = \text{diag}\left(\omega_\mu^2, \omega_\alpha^2\right)$, α_t is the time-varying loading coefficient that controls the effects of $\exp(h_t)$ on the inflation rate, y_t, and $y_{1:T} = (y_1, \ldots, y_T)'$. Gibbs sampling estimation of (2.1)–(2.3) is nontrivial since h_t appears in both the conditional mean and the conditional variance. Thus, well-known methods such as Kim et al. (1998) cannot be used to draw $h_{1:T} \sim p(h_{1:T} \mid \theta, \gamma_{1:T}, y_{1:T})$. To overcome this challenge, one can implement an accept–reject Metropolis–Hastings (AR-MH) procedure to draw $h_{1:T} \sim p(h_{1:T} \mid \theta, \gamma_{1:T}, y_{1:T})$, see Chan (2014). Conditional on $h_{1:T}$, drawing $\gamma_{1:T} \sim p(\gamma_{1:T} \mid \theta, h_{1:T}, y_{1:T})$ and $\theta \sim p(\theta \mid h_{1:T}, \gamma_{1:T}, y_{1:T})$, where $\theta = \left(\mu_h, \phi_h, \sigma_h^2, \omega_\mu^2, \omega_\alpha^2\right)'$ is easy using standard Gibbs sampling techniques. However, most Gibbs sampling approaches face two common difficulties. They are:

- (a) Obtaining the integrated likelihood at each Gibbs iteration for model $\mathcal{M}_k$, $p\left(y_{1:T} \mid \theta^{(i)}, \mathcal{M}_k\right)$, $i = 1, \ldots, N$, which is essential for ML and DIC computation is cumbersome and besides the main Gibbs algorithm requires additional coding effort for each separate model specification.
- (b) We cannot use a unified sampling algorithm. For instance, assume that we want to compare (2.1)–(2.3) with $y_t = \mu_t + \varepsilon_t^y$, where $\varepsilon_t^y \sim N(0, \exp(h_t))$ and $\mu_t = \mu_{t-1} + \varepsilon_t^\mu$, where $\varepsilon_t^\mu \sim N\left(0, \omega_\mu^2\right)$. In this case, we can (I) Augment $p(\theta, \mu_{1:T}, h_{1:T} \mid y_{1:T})$ to include the mixture component indicators of Kim et al. (1998), $z_{1:T}$. Thus, we also sample $z_{1:T} \sim p\left(z_{1:T} \mid \theta, \mu_{1:T}, h_{1:T}, y_{1:T}\right)$, which requires programming a new routine, or (II) Change (among other things) the first and second order derivatives of $p(y_{1:T} \mid \theta, \gamma_{1:T}, h_{1:T})$ in the AR-MH procedure of Chan (2014).

3. Estimation

Intuitively, PG-AS combines Gibbs sampling and sequential Monte Carlo (SMC) methods.[1] In PG-AS, we act as if we are operating within a Gibbs sampling setting except for one difference, namely, that we draw $x_{1:T} \sim p(x_{1:T} \mid \theta, y_{1:T})$, where $x_t = (h_t, \gamma_t)'$ (all-at-once) using the conditional particle filter with ancestor sampling, CPF-AS, see Appendix or Lindsten et al. (2014). Thus, by using PG-AS, we can automatically reduce the number of sampling steps. Conditional on $x_{1:T}$, we sample $\theta \sim p(\theta \mid x_{1:T}, y_{1:T})$ element-by-element using standard Gibbs techniques. Furthermore, $p\left(y_{1:T} \mid \theta^{(i)}, \mathcal{M}_k\right)$, $i = 1, \ldots, N$ is automatically available as a byproduct of CPF-AS. Therefore, we do not need to perform any modifications or program a new routine to obtain it. On the other hand, if we were to obtain $p\left(y_{1:T} \mid \theta^{(i)}, \mathcal{M}_k\right)$ within a Gibbs sampling approach, then besides sampling $x_{1:T}^{(i)}$ and $\theta^{(i)}$, we would for example need to run a separate particle filter (for each model) at each Gibbs iteration. Adding this step increases the computational and coding burden.

For each model, we use $p\left(y_{1:T} \mid \theta^{(i)}, \mathcal{M}_k\right)$, $i = 1, \ldots, N$, to compute ML and DIC. Specifically, we calculate ML using the method of Gelfand–Dey, see Koop (2003). DIC is calculated as $D\left(\bar{\theta}\right) + 2p_D$, where $p_D = \overline{D(\theta)} - D\left(\bar{\theta}\right)$. This parameter describes the complexity of the model, serving as a penalization term that corrects deviance's propensity towards models with more parameters. We estimate $\overline{D(\theta)}$ using $N^{-1}\Sigma_{i=1}^N - 2\log p\left(y_{1:T} \mid \theta^{(i)}, \mathcal{M}_k\right)$ and $D\left(\bar{\theta}\right) = -2\log p\left(y_{1:T} \mid \bar{\theta}, \mathcal{M}_k\right)$. $\bar{\theta}$ is the mean or mode of $\{\theta^{(i)}\}_{i=1}^N$. The best model is the one with the highest (smallest) ML (DIC). Note that contrary to ML, DIC is considered as a measure of model fit plus a degree of complexity rather than solely a goodness of fit measure. In other words: ML addresses how well our prior predicts the data, whereas DIC addresses how well the posterior might predict future observations generated by the same parameters that give rise to the observed data.

Finally, PG-AS requires limited design requirements from the practitioner's part, especially when one desires to change some features in a particular model. For instance, in order to estimate (2.1)–(2.3) without SVM effects, we can easily skip drawing particles for α_t, omit sampling ω_α^2 and then change (2.1) inside CPF-AS to $y_t = \mu_t + \varepsilon_t^y$. We sample $\tilde{x}_{1:T} \sim p\left(\tilde{x}_{1:T} \mid \tilde{\theta}, y_{1:T}\right)$, where $\tilde{x}_t = (\mu_t, h_t)'$ all-at-once and then proceed to sample $\tilde{\theta} \sim p\left(\tilde{\theta} \mid \tilde{x}_{1:T}, y_{1:T}\right)$. Furthermore, we automatically have $p\left(y_{1:T} \mid \tilde{\theta}^{(i)}\right)$, which we use to compute ML and DIC.

4. Results

Our data consists of US and UK quarterly seasonally adjusted CPI inflation rates from 1947q1 and 1955q1 to 2014q4, respectively.[2] Our models are summarized in Table 1. For θ, we choose the same priors as Chan (2013, 2014). We set the number of particles, M, to 100 and take 20 000 draws from $p(\theta, \gamma_{1:T}, h_{1:T} \mid y_{1:T})$ after a burn-in of 5000. The estimated MLs and DICs are reported in Table 2. In order to calculate the numerical standard errors (NSE) of ML and DIC, we use "computational force", which is simply reproducing our calculations some 20 times and estimating NSEs by their sample standard deviations, see also Berg et al. (2004).

Overall, we obtain very interesting results. For instance, we see that UC-SVM ($\mathcal{M}_6$) performs best, regardless of criteria. However, compared to US data, incorporating time-variation in α leads to relatively less improvements over $\mathcal{M}_5$ for the UK data. The Bayes factor of $\mathcal{M}_6$ over $\mathcal{M}_5$, $BF_{\mathcal{M}_6,\mathcal{M}_5}$, is $\exp(5.28)$ compared to $\exp(16.26)$ for the US data. Furthermore, the difference in DIC between $\mathcal{M}_6$ and $\mathcal{M}_5$ is 0.14 for the UK data. This indicates that adding the additional complexity, i.e. time-variation in α does not lead to any improvements in terms of model fit according to DIC, underlying the difference between ML and DIC measures. For both series, compared to $\mathcal{M}_2$, incorporating SV effects in the core component, adding moving average errors and SVM effects all lead to improvements, both in terms of ML and DIC. The NSEs show that ML and DIC are accurately estimated.

We collect the final output and report estimation results for $\mathcal{M}_6$ in Table 3. In general, parameter estimates are similar to those

[1] PG-AS has shown to be very robust to path degeneracy problems, which other techniques like for instance the particle Gibbs (PG) sampler of Andrieu et al. (2010) can encounter, see Lindsten et al. (2014) for more details.

[2] Both data series are downloadable from FRED's website. We seasonally adjust the UK series using the X-12-ARIMA procedure.