Second homework assignment

PhD in Business Economics | Course: Advanced Bayesian Econometrics
Professor: Hedibert Freitas Lopes | Due date: 12h, February 11th, 2021.

Use, preferably, `Rmarkdown` (via `RStudio`) to produce your report in PDF or HTML.

Fitting Gaussian and Student's t ARMA(1,1) model

Let us assume that some observed time series data $\{y_1, \ldots, y_n\}$ follows an ARMA(1,1) model

$$y_t = \phi y_{t-1} + \varepsilon_t + \gamma \varepsilon_{t-1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. either $\mathcal{M}_0 : N(0, \sigma^2)$ or $\mathcal{M}_1 : t_\nu(0, \tau^2)$, where $\tau^2 = (\nu - 2)/\nu\sigma^2$. We will keep ν fixed and known throughout. In addition, in order to simplify the homework, we will assume that $y_0 = \varepsilon_0 = 0$. Therefore, it is easy to see that $\{\varepsilon_1, \ldots, \varepsilon_n\}$ are deterministically obtained from $\theta = (\phi, \gamma, \sigma)$ and the data $y^n = \{y_1, \ldots, y_n\}$: $\varepsilon_1 = y_1$ and $\varepsilon_t = y_t - \phi y_{t-1} - \gamma \varepsilon_{t-1}$, for $t = 2, \ldots, n$.

Likelihood functions. To avoid overloading the notation, let us drop (ε_0, y_0) in what follows. The likelihood functions are, therefore,

$$\begin{aligned}
\mathcal{L}(\theta | y^n, \mathcal{M}_0) &= (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_{t=1}^n \varepsilon_t^2}{2\sigma^2} \right\} \\
\mathcal{L}(\theta | y^n, \mathcal{M}_1) &= \left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\tau^2}} \right)^n \prod_{t=1}^n \left(1 + \frac{1}{\nu} \frac{\varepsilon_t^2}{\tau^2} \right)^{-\frac{\nu+1}{2}}.
\end{aligned}$$

Prior distribution. Let us assume that

$$p(\theta) = p(\phi, \gamma, \sigma^2) = p(\phi)p(\gamma)p(\sigma^2),$$

for

$$\begin{aligned}
\phi &\sim U(-1, 1) \\
\gamma &\sim U(-1, 1) \\
\sigma^2 &\sim IG(5/2, 5(1.4)/2).
\end{aligned}$$

Hence, we are constraining our inference to the class of stationary and invertible ARMA(1,1) models. In additional, prior mean, mode and standard deviation for σ^2 is around 2.33, 1 and 3.3, respectively. Also, $Pr(\sigma^2 \in (0.33, 26.3)) \approx 99.9\%$, so $\sigma < 0.5$ or $\sigma > 5$ are essentially ruled out as well.

Simulating some data. You should simulate two datasets of size $n = 400$, one with Gaussian errors and the other with Student's t errors where $\sigma = 1$, $\nu = 4$, $\phi = 0.98$ and $\gamma = -0.64$. Feel free to use the following R script:

```
set.seed(12345)
n =400
sig=1.0
nu=4
phi=0.98
theta=-0.64
tau = sqrt((nu-2)/nu)*sig
e.n = sig*rnorm(n)
e.t  = tau*rt(n,df=nu)
y.n = rep(0,n)
y.t = rep(0,n)
y.n[1] = e.n[1]
y.t[1] = e.t[1]
for (t in 2:n){
  y.n[t] = phi*y.n[t-1]+e.n[t]+theta*e.n[t-1]
  y.t[t] = phi*y.t[t-1]+e.t[t]+theta*e.t[t-1]
}
par(mfrow=c(1,1))
ts.plot(cbind(y.n,y.t),col=1:2,main="ARMA(1,1) data")
legend("bottomleft",legend=c("Gaussian","Student's t"),col=1:2,lty=1,bty="n")
```

Questions: Answer the following questions for each one of the two datasets generated by the previous script.

1. **Maximum likelihood inference.**
   What are the maximum likelihood estimates (MLE) of θ under both models? Use the `R` function `nlm` to minimize the negative of the likelihood functions. Are the results similar to the ones from the `R` function `arima(y,order=c(1,0,1))`?

2. **Bayesian inference via Monte Carlo methods.**

   (a) Use sampling importance resampling (SIR) to sample from both posterior distributions of θ:

   $$\begin{aligned} p(\theta|y^n, \mathcal{M}_0) &\propto \mathcal{L}(\theta|y^n, \mathcal{M}_0)p(\theta) \\ p(\theta|y^n, \mathcal{M}_1) &\propto \mathcal{L}(\theta|y^n, \mathcal{M}_1)p(\theta). \end{aligned}$$

   Use these draws whenever necessary in the next several questions.

   (b) Compute posterior means, medians and 95% credibility interval for ϕ, γ and σ^2. Are posterior means (and medians) similar to their MLE counterparts?

(c) Plot the contours of the posterior density. What are the posterior probabilities that $\phi > 0.9$ under both models?

3. **Prior predictive, Bayes factor and posterior model probability.**

(a) Compute both prior predictive $p(y^n|\mathcal{M}_0)$ and $p(y^n|\mathcal{M}_1)$. We can approximate the prior predictive densities, for $j = 0, 1$

$$p(y^n|\mathcal{M}_j) = \int_0^\infty \int_{-1}^1 \int_{-1}^1 \prod_{t=1}^n p(y_t|y_{t-1}, \theta, \mathcal{M}_j) d\phi d\gamma d\sigma^2,$$

via Monte Carlo by

$$\widehat{p}(y^n|\mathcal{M}_j) = \frac{1}{M} \sum_{i=1}^M \prod_{t=1}^n p(y_t|y_{t-1}, \phi^{(i)}, \gamma^{(i)}, \sigma^{2(i)} \mathcal{M}_j),$$

where $\theta^{(1)}, \ldots, \theta^{(M)}$ are draws from the prior $p(\theta)$. Let us use $M = 100,000$.

(b) Compute a MC approximation to the Bayes factor:

$$B_{01} = \frac{p(y^n|\mathcal{M}_0)}{p(y^n|\mathcal{M}_1)}.$$

(c) Finally, the posterior model odds can be computed as

$$\frac{Pr(\mathcal{M}_0|y^n)}{Pr(\mathcal{M}_1|y^n)} = \frac{Pr(\mathcal{M}_0)}{Pr(\mathcal{M}_1)} \times B_{01},$$

where $Pr(\mathcal{M}_0)$ and $Pr(\mathcal{M}_1)$ are the prior probabilities assigned to models $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively. Assuming $Pr(\mathcal{M}_0) = Pr(\mathcal{M}_1)$, obtain a MC approximation to $Pr(\mathcal{M}_0|y^n)$, the posterior model probability of the Gaussian model.

Discuss your findings.

Hints and clarification: Below are a few hints about specific issues raised by some of you over the last couple of days while handling this homework assignment.

[A] Gamma, inverse-gamma and their parameterizations.

The notation $\sigma^2 \sim IG(5/2, 5(1.4)/2)$ means that σ^2 is distributed as an inverse-gamma with parameters $a = 5/2 = 2.5$ and $b = 5(1.4)/2 = 3.5$. Why not expressing the prior directly as $\sigma^2 \sim IG(2.5, 3.5)$? I obviously could have done that, but I like the $\sigma^2 \sim IG(\nu/2, \nu\tau^2/2)$, so ν plays a role similar to the number of degrees of freedom and τ^2 resembles the prior mean of σ^2. Recall that, when $X \sim IG(a, b)$,

$$
\begin{aligned}
E(X) &= \frac{b}{a-1}, \quad a > 1, \\
V(X) &= \frac{b^2}{(a-1)^2(a-2)}, \quad a > 2, \\
p(x) &= \frac{b^a}{\Gamma(a)} x^{-(a+1)} \exp\{-b/x\},
\end{aligned}
$$

where $\Gamma(a)$ is the gamma function ($\Gamma(a) = (a-1)\Gamma(a-1)$). Another property is that when $X \sim IG(a, b)$, t follows that $Y = 1/X \sim G(a, b)$, where $E(Y) = a/b$, $V(Y) = a/b^2$ and $p(y) = \frac{b^a}{\Gamma(a)} y^{(a-1)} \exp\{-by\}$. Notice that $E(X) = b/(a-1) \neq b/a = 1/E(Y)$, unless a and b are both very large. With the (ν, τ^2) notation, that translates into ν being very large (or very large degrees of freedom!)

The lesson here is that only the normal distribution has parameters that are explicitly its mean ad its variance. All other distributions (Binomial, Poisson, Gamma, Beta, Hypergeometric, Weibull, Gumbel, Student's t, etc) have means and variances which are functions of its parameters.

[B] Outline of a generic SIR

There is a lot of confusion here about what is drawn from the proposal distribution and what is to evaluate a function at a specific value. Let us consider a um parameter, θ, implementation of SIR where the target distribution is $\pi(\theta)$. In our most common context, $\pi(\theta) \propto p(\theta)p(y|\theta)$ where $p(\theta)$ is the prior of θ and $p(y|\theta)$ is the likelihood of θ. What we would do in order to draw from $\pi(\theta)$ via SIR?

1. Pick a proposal $q(\theta)$ that is easy to sample from, easy to evaluate and that resembles $\pi(\theta)$.
2. Draw $\{\widetilde{\theta}^{(1)}, \ldots, \widetilde{\theta}^{(M)}\}$ from the proposal $q(\theta)$.
3. Evaluate the resampling weights

$$
w^{(i)} = \frac{\pi(\widetilde{\theta}^{(i)})}{q(\widetilde{\theta}^{(i)})}, \quad i = 1, \ldots, M.
$$

   Notice that we are evaluating both $\pi(\cdot)$ and $q(\cdot)$ at the sampled draws. We are not evaluating on any pre-specified grid.

4. Sample from the discrete set $\{\widetilde{\theta}^{(1)}, \ldots, \widetilde{\theta}^{(M)}\}$ with sampling weights $\{w^{(1)}, \ldots, w^{(M)}\}$.

5. Let $\{\theta^{(1)}, \ldots, \theta^{(M)}\}$ be the resampled draws.
6. The histogram of $\{\theta^{(1)}, \ldots, \theta^{(M)}\}$ should approximate $\pi(\theta)$ when M is considerably large.
7. Similarly, $E(g(\theta)) = \int g(\theta)\pi(\theta)d\theta$ is approximated by $\sum_{i=1}^{M} g(\theta^{(j)})/M$.

In the problem at hand, $\theta = (\phi, \gamma, \sigma^2)$ and $\pi(\theta) \propto \mathcal{L}(\theta|y^n)p(\theta)$. Many people are choosing $q(\theta)$ as $p(\theta)$, the prior of θ. That is fine and it is a common practice, which leads to weights proportional to the likelihood $\mathcal{L}(\cdot|y^n)$ evaluated at the draws from the prior. The main problem with this approach appears when the prior is too flat relative to the likelihood. In this case, the vast majority of the proposed draws will have weights virtually and practically equal to zero and only a few proposed draws will have weights very large.

Finally, regardless of how you choose your proposal density $q(\theta)$, it is usually advisable to consider logarithms of the prior, of the likelihood and of the proposal (therefore, of the weights as well) to avoid underflow and/or overflow during your computations. Try the following piece of code and check the weights when drawing from a $N(0, \sigma^2_{prop})$ to obtain draws from a $N(0, 1)$. Notice some weights as small as e^{-238}. These ratios can easily collapse at zero, Inf or NA. Play around with σ_{prop} and see how easy is to crash this code!

Now, if you want to check if the SIR draws are actually draws from the target density, i.e. the N(0,1), you simple add its curve on the top of the histogram. More specifically, you simple add a line of code `lines(thetas,dnorm(thetas),col=2)`, where `thetas` is a grid of points say from -5 to 5, or `thetas=seq(-5,5,length=100)`.

```
set.seed(12345)
M        = 100000
sig.prop = 10
draws = rnorm(M,0,sig.prop)
w = dnorm(draws)/dnorm(draws,0,sig.prop)
w1 = dnorm(draws,log=TRUE)-dnorm(draws,0,sig.prop,log=TRUE)
w2 = exp(w1-max(w1))
sort(w)[c(1:3,(M-2):M)]
sort(w1)[c(1:3,(M-2):M)]
sort(w2)[c(1:3,(M-2):M)]
draws1 = sample(draws,replace=TRUE,size=M/2,prob=w2)
par(mfrow=c(1,2))
hist(draws,prob=TRUE,xlab="",main=paste("Draws from N(0,",sig.prop,"^2)",sep=""))
hist(draws1,prob=TRUE,xlab="",main="Draws from N(0,1)")
thetas=seq(-5,5,length=100)
lines(thetas,dnorm(thetas),col=2,lwd=2)

# results
#[1]  0  0  0 10 10 10
#[1] -1540.618531 -1028.417719 -1026.298068     2.302585     2.302585     2.302585
#[1] 0 0 0 1 1 1
```