
Third homework assignment

PhD in Business Economics
Professor: Hedibert Freitas Lopes

Course: Econometrics III
Due date: 9am, February 23rd, 2020.

You can work individually or in pairs.

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

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

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

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

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. either $\mathcal{M}_0 : N(0, \sigma^2)$ or $\mathcal{M}_1 : t_\nu(0, \tau^2)$, where $\tau^2 = (\nu - 2)/\nu\sigma^2$. Both σ^2 and ν are going to be kept 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 $\{\varepsilon_1, \dots, \varepsilon_n\}$ are deterministically obtained from the parameters $\theta = (\phi, \gamma)$ and the data $y^n = \{y_1, \dots, y_n\}$. More specifically, $\varepsilon_1 = y_1$ and $\varepsilon_t = y_t - \phi y_{t-1} - \gamma \varepsilon_{t-1}$, for $t = 2, \dots, n$.

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

$$\mathcal{L}(\theta|y^n, \mathcal{M}_0) = c_0 \exp\left\{-\frac{\sum_{t=1}^n \varepsilon_t^2}{2\sigma^2}\right\} \quad \text{and} \quad \mathcal{L}(\theta|y^n, \mathcal{M}_1) = c_1 \prod_{t=1}^n \left(1 + \frac{1}{\nu} \frac{\varepsilon_t^2}{\tau^2}\right)^{-\frac{\nu+1}{2}},$$

where the normalizing constants are $c_0 = (2\pi\sigma^2)^{-n/2}$ and $c_1 = \left(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\tau^2}}\right)^n$ are irrelevant for posterior inference, but very important when computing predictive densities.

Hint: Once you have computed the errors $\varepsilon_1, \dots, \varepsilon_n$ conditional on ϕ and γ , the above likelihoods (their logs is better for computation!) are easily computed in R as

```
loglike.gaussianmodel = sum(dnorm(eps, , sig, log=TRUE))
loglike.studenttmodel = sum(dt(eps/tau, df=nu, log=TRUE)) - n/log(tau)
```

Prior distribution. Let us assume that $p(\theta) = p(\phi)p(\gamma)$, for $\phi \sim U(-1, 1)$ and $\gamma \sim U(-1, 1)$, so we are only entertaining stationary and invertible ARMA(1,1) models.

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
gamma =-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]+gamma*e.n[t-1]
  y.t[t] = phi*y.t[t-1]+e.t[t]+gamma*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 errors ", "Student's t(4) errors"),col=1:2,lty=1,bty="n")
```

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

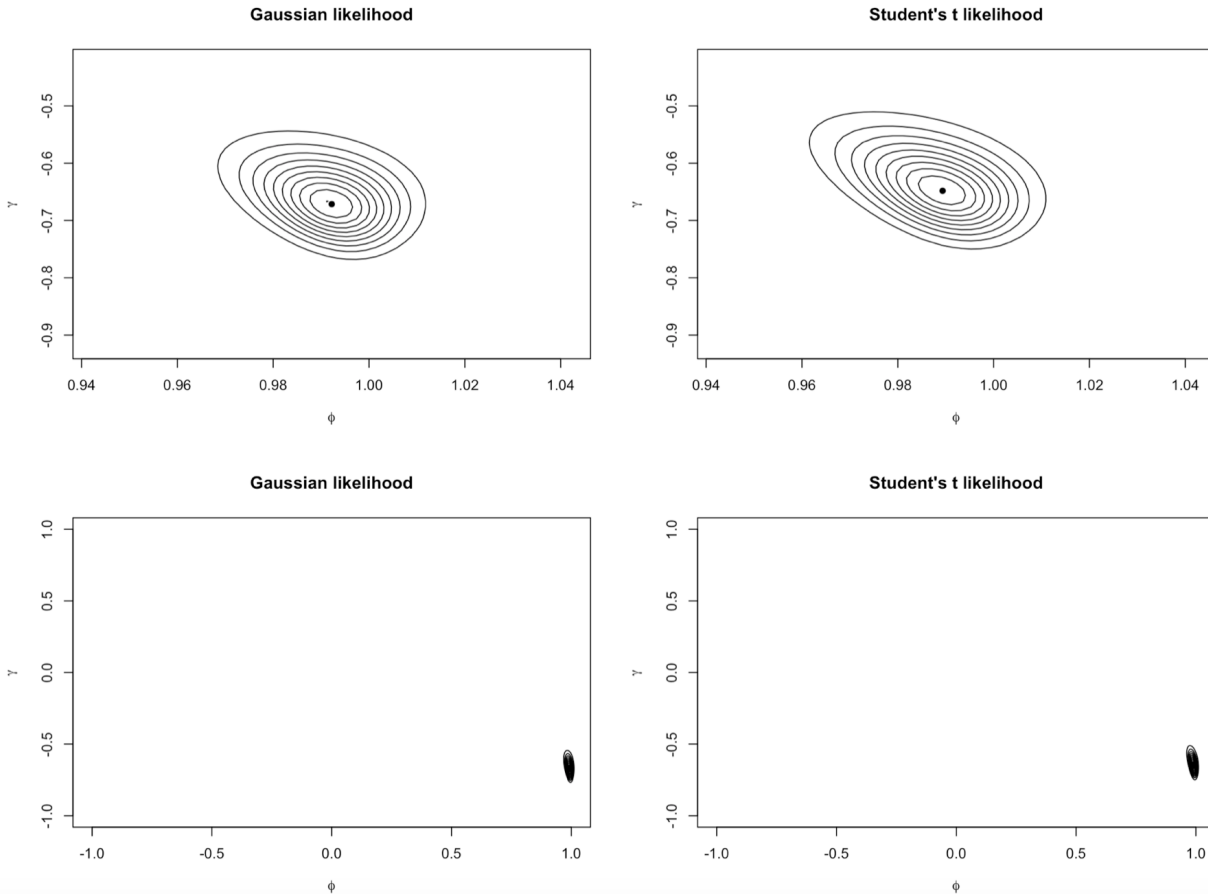
1. What are the maximum likelihood estimates (MLE) of (ϕ, θ) under both models? Use the R function `nlm` to minimize the negative of the likelihood functions. Are the results similar to the ones from the R function `arima(y,order=c(1,0,1))`? Take a look at the hint below.
2. Use sampling importance resampling (SIR) to sample from both posteriors of θ . Recall that the log-posterior densities are

$$\log p(\theta|y^n, \mathcal{M}_0) = \log c_0 - 0.5 \sum_{t=1}^n \varepsilon_t^2 / \sigma^2$$

$$\log p(\theta|y^n, \mathcal{M}_1) = \log c_1 - 0.5(\nu + 1) \sum_{t=1}^n \log \left(1 + \frac{1}{\nu} \frac{\varepsilon_t^2}{\tau^2} \right).$$

Use these draws whenever necessary in the next several questions.

Hint: Observe where $p(\theta|y^n, \mathcal{M}_0)$ and $p(\theta|y^n, \mathcal{M}_1)$ in the domain of the θ , i.e. the square $(-1, 1) \times (-1, 1)$. In fact, there they are in the tiny little lower right corner of the square. Therefore, using the prior of θ might not be a wise idea since virtually all draws will be rejected and your resample will contain only a handful of those prior draws. Try using a tighter proposal, say a uniform in the tinier square $(0.94, 1.04) \times (-0.9, -0.5)$ (as in the top row of the figure).



3. Compute posterior means, medians and 95% credibility interval for ϕ and γ . Are posterior means (and medians) similar to their MLE counterparts?
4. Plot the contours of the posterior density. What are the posterior probabilities that $\gamma > 1.0$ under both Gaussian and Student's t models?
5. Compute both predictive values $p(y^n|\mathcal{M}_0)$ and $p(y^n|\mathcal{M}_1)$. Since $p(\theta) = 1$ for $\theta \in \mathcal{A} = (-1, 1) \times (-1, 1)$, we can Monte Carlo integration to approximate the **prior predictive** densities, for $j = 0, 1$,

$$p(y^n|\mathcal{M}_j) = \int_{\mathcal{A}} \frac{\prod_{t=1}^n p(y_t|y_{t-1}, \theta, \mathcal{M}_j)}{q(\theta)} q(\theta) d\theta,$$

by

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

where $\theta^{(1)}, \dots, \theta^{(M)}$ are draws from the proposal, $q(\theta) = 1/\xi$, a uniform distribution on the square $\mathcal{B} = (0.94, 1.04) \times (-0.9, -0.5)$. From the previous hint, we see that all the parameter region in \mathcal{A} that is outside the tiny little region \mathcal{B} has negligible posterior density, i.e. $p(\theta | y^n) = 0$ for $\theta \in \mathcal{A} \setminus \mathcal{B}$, hence need not be considered for Monte Carlo approximation. Here ξ is the area under \mathcal{B} , i.e. $\xi = (1.04 - 0.94)(-0.5 - (-0.9)) = 0.04$. I suggest using at least $M = 100,000$ draws from $q(\theta)$.

Let us also compute the Bayes factor of the Gaussian model against the Student's t model

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

and discuss your findings.

6. Compute $E(y_{n+1} | y^n, \mathcal{M}_j)$ and $V(y_{n+1} | y^n, \mathcal{M}_j)$ for $j = 0, 1$. Similar to the previous question, it can be shown that

$$E[g(y_{n+1}) | y^n, \mathcal{M}_j] = \int_{\mathcal{A}} E[g(y_{n+1}) | y^n, \theta, \mathcal{M}_j] p(\theta | y^n) d\theta,$$

which is an integral similar to the one found when computing $p(y^n | \mathcal{M}_j)$ in 5) above. The difference is that the integral is with respect to the posterior of θ , while in 5 it was with respect to the prior. These new quantities are from the **posterior predictive** densities. We already have draws from these posterior distributions from 2, 3 and 4 above). We have now all ingredients necessary to compute $E(y_{n+1} | y^n, \mathcal{M}_j)$ and $V(y_{n+1} | y^n, \mathcal{M}_j)$ for $j = 0, 1$ and only need to consider $g(y_{n+1}) = y_{n+1}$ and $g(y_{n+1}) = y_{n+1}^2$.

2 Threshold autoregressive (TAR) model

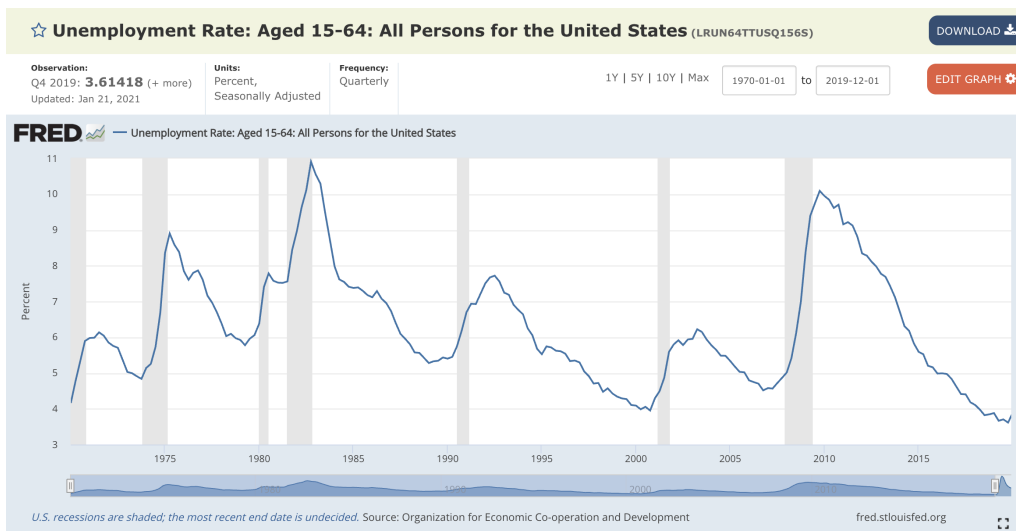
In the third edition of his *Analysis of Financial Time Series* book, Ruey Tsay introduces the 2-regime threshold autoregressive model of order p , or the TAR(p) model:

$$y_t | y_{t-1}, \dots, y_{t-p}, \theta \sim \begin{cases} N \left(\phi_0 + \sum_{j=1}^p \phi_j y_{t-j}, \sigma^2 \right) & \text{if } y_{t-1} < \gamma \\ N \left(\beta_0 + \sum_{j=1}^p \beta_j y_{t-j}, \tau^2 \right) & \text{if } y_{t-1} \geq \gamma \end{cases}$$

for $t = 1, \dots, n$ and $\theta = (\phi, \sigma, \beta, \tau, \gamma)$ of dimension $k = 2p + 5$. See Section 4.1.2 (pages 179-182) and Section 4.5 (pages 218-222), for further details.

Quarterly U.S. civilian unemployment rate, seasonally adjusted. Section 4.5 (Application) of Ruey's book starts by illustrating the estimation of nonlinear time series models by analyzing the quarterly U.S. civilian unemployment rate, seasonally adjusted, from 1948Q1 to 1993Q4. Equation (4.53) shows a fitted TAR(2) model to $\Delta y_t = (y_t - y_{t-1})$, the first difference of the unemployment rate. He obtained, by maximum likelihood estimation (MLE), $\hat{\phi} = (0.01, 0.73, 0.10)$, $\hat{\beta} = (0.18, 0.80, -0.56)$, $\hat{\sigma} = 0.76$, $\hat{\tau} = 0.165$ and $\hat{\gamma} = 0.1$, and standard errors $se(\hat{\phi}) = (0.03, 0.10, 0.12)$ and $se(\hat{\beta}) = (0.09, 0.1, 0.16)$.

Updating the data. Your task is three-fold. Firstly, you should use an updated version of the time series. Downloaded the time series from <https://fred.stlouisfed.org/series/LRUN64TTUSQ156S> and let us set the period between 1970Q1 to 2019Q4. Secondly, you should obtain MCMC-based posterior inference for θ in the above 2-regime Gaussian TAR(2) model using the data up to 2017Q4. Thirdly, you should use the last 8 quarters (2018Q1-2019.Q4) to perform out-of-sample forecasting. The following plot was taken from FRED, Federal Reserve Bank of St. Louis, webpage.



Prior. Let us use conditionally conjugate independent priors for the 3-dimensional vectors of parameters ϕ and β , i.e. $N(0, I_3)$ is the prior distribution for both ϕ and β . Also, let us use a Half-Cauchy(0, 1) distribution for both standard deviation parameters σ and τ . For the threshold parameter γ , let its prior be a uniform distribution from y_{min} to y_{max} , i.e. between the minimum and the maximum of the observed data.

Posterior. It should come at no surprise that the posterior of θ does not resemble any known distribution. More specifically,

$$p(\phi, \sigma, \beta, \tau, \gamma | y_{1:n}) \propto \left\{ \frac{\exp\{-0.5(\phi'\phi)\}}{1 + \sigma^2} \prod_{\mathcal{A}_\gamma} p_N(y_t, x'_t \phi, \sigma^2) \right\} \left\{ \frac{\exp\{-0.5(\beta'\beta)\}}{1 + \tau^2} \prod_{\mathcal{B}_\gamma} p_N(y_t, x'_t \beta, \tau^2) \right\}, \quad (1)$$

where $y_0 = y_{-1} = y_{-2} = 0$, $x'_t = (1, y_{t-1}, y_{t-2})$, $\mathcal{A}_\gamma = \{t = 1, \dots, n \mid y_{t-1} < \gamma\}$ and $\mathcal{B}_\gamma = \{t = 1, \dots, n \mid y_{t-1} \geq \gamma\}$. Hence, posterior inference should be based on some sort of MCMC scheme. Also, it should be obvious by now that ϕ and β are conditionally independent given γ (the same applies σ and τ). Notice that (ϕ, σ) and (β, τ) , conditionally on γ , are completely separated in Equation (1).

MCMC algorithm. You will notice that the full posterior distributions of ϕ and β

$$p(\phi | y_1, \dots, y_n, \sigma, \phi, \tau, \gamma) \quad \text{and} \quad p(\beta | y_1, \dots, y_n, \tau, \phi, \sigma, \gamma),$$

are both Gaussian, therefore Gibbs steps are available. The same cannot be said about the full conditional distributions of the three remaining parameters, i.e. σ , τ and γ . I suggest you use Gaussian random-walk Metropolis-Hastings (RW-MH) proposals for $\log \sigma$, $\log \tau$ and $\log \gamma$, with the same tuning variance for all three parameters, say $\nu^2 = 0.01$. You should be extra careful when computing the RW-MH acceptance probabilities as it should include the Jacobians of the logarithm transformations, i.e. when

$$Y \sim p_Y(y) \quad \text{and} \quad Y = g(X), \quad \text{then} \quad p_X(x) = p_Y(g(y)) |g'(y)|.$$

Therefore, our MCMC scheme has two Gibbs steps, each sampling three parameters, and three RW-MH steps, each one sampling only one parameter. Set the initial state of the Markov chain at $\theta^{(0)}$ (suggestion: use the MLE $\hat{\theta}$). Then, algorithmically, the MCMC scheme runs as follows:

For $i = 0, \dots, M - 1$,

- [ϕ] Sample $\phi^{(i+1)}$ from $p(\phi | y_{1:n}, \sigma, \gamma)$ - Gibbs step;
- [β] Sample $\beta^{(i+1)}$ from $p(\beta | y_{1:n}, \tau, \gamma)$ - Gibbs step;
- [σ] Sample $\log \sigma^*$ from $N(\log \sigma^{(i)}, \nu^2)$ - RW-MH step:

$$\sigma^{(i+1)} = \begin{cases} \sigma^* & \text{with probability } \alpha \\ \sigma^{(i)} & \text{with probability } 1 - \alpha \end{cases}$$

where

$$\alpha = \min \left\{ 1, \frac{p(\phi^{(i)}, \sigma^*, \beta^{(i)}, \tau^{(i)}, \gamma^{(i)} | y_{1:n})}{p(\phi^{(i)}, \sigma^{(i)}, \beta^{(i)}, \tau^{(i)}, \gamma^{(i)} | y_{1:n})} \times \frac{\frac{1}{\sigma^{(i)}}}{\frac{1}{\sigma^*}} \right\}$$

- [τ] Sample $\log \tau^*$ from $N(\log \tau^{(i)}, \nu^2)$ - RW-MH step (the derivation is similar to the one for σ);
- [γ] Sample $\log \gamma^*$ from $N(\log \gamma^{(i)}, \nu^2)$ - RW-MH step (the derivation is similar to the one for σ);

Review: 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 $\{\tilde{\theta}^{(1)}, \dots, \tilde{\theta}^{(M)}\}$ from the proposal $q(\theta)$.
3. Evaluate the resampling weights

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

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

In the first question (ARMA(1,1) model), $\theta = (\phi, \gamma)$ and $\pi(\theta) \propto \mathcal{L}(\theta|y^n)p(\theta)$. A natural choice for $q(\theta)$ is the prior $p(\theta)$. 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 $N(0, \sigma_{prop}^2)$ 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

```