NOTES ON REGRESSION WITH t-DISTRIBUTED ERRORS

The R code posted with the exercise may not be easy to interpret. Here is the math the code implements.

The model is

$$y_t = X_t\beta + \varepsilon_t \tag{1}$$

$$\frac{\varepsilon_t}{\sigma} \mid X \sim t(\nu)\,, \tag{2}$$

where ν is the degrees of freedom parameter for the t distribution. The likelihood for the sample is therefore, using $\tau(\cdot;\nu)$ to denote the pdf of the $t(\nu)$ distribution,

$$\sigma^{-T}\prod_{t=1}^{T}\tau\left(\frac{y_t - X_t\beta}{\sigma};\nu\right)\,.$$

If we want to do model comparison, we need to use a proper prior on β and σ, but if we are only comparing models with differing values of the explicit parameters β, σ, and ν, we can use the likelihood as the posterior (i.e. treat the prior as flat in the relevant range). [Note, thugh, that as we discussed in class a flat prior on the degrees of freedom itself produces a non-integrable likelihood, which therefore cannot be used for MCMC evaluation of the posterior when we sample ν along with other parameters. A flat prior on $1/\nu$ over its range of $(0,1)$, though, is equivalent to a prior pdf of $1/\nu^2$ on $(1,\infty)$ for ν itself, which is integrable.]

This likelihood can be calculated in a single line of R or matlab code, and one could use it to implement random walk Metropolis posterior simulation for the joint posterior distribution of β, σ for fixed ν, or jointly, sampling ν as well as the other parameters. You might try this, to see if it works better than the approach below, but this is not required.

A random variable that has a $t(\nu)$ distribution has the same distribution as $z/\sqrt{w/\delta}$, where $z \sim N(0,1)$, $w \sim \chi^2(\delta)$, and z and w are independent. In other words, a $t(\nu)$ random variable can be thought of as a normal random variable with a randomly chosen variance. This lets us introduce an auxiliary sequence of random variables, $\{w_t\}$ to denote the inverses of the random variances of the residuals $\{\varepsilon_t/\sigma\}$. This idea, of introducing additional unobserved random variables to make MCMC sampling easier to implement is known at **data augmentation**. Then instead of using the t pdf itself, we can write the joint density of y and w conditional on β, ν and σ^2 as

$$\Gamma(\nu/2)^{-T}\,2^{-T\nu/2}\prod_{t=1}^{T}\left(w_t^{\frac{\nu}{2}-1}\right)e^{-\frac{1}{2}\sum_1^T w_t}\sigma^{-T}(2\pi)^{-T/2}e^{-\sum_1^T w_t(y_t-X_t\beta)^2/(2\nu\sigma^2)}\,. \tag{3}$$

With σ, ν and $\{w_t\}$ held constant, this likelihood behaves, as a function of β, like a normal pdf with mean the weighted least-squares estimate, where the $\{w_t\}$ are the weights on the cross products, and covariance matrix $\sigma^2(\sum w_t X_t' X_t)^{-1}$. We know how to calculate these objects and hence how to generate a draw from this normal distribution.

With $\{w_t\}$, ν and β fixed, the likelihood as a function of σ is proportional to

$$\sigma^{-T} e^{-\sum \hat{\varepsilon}_t^2/(2\nu\sigma^2)}, \tag{4}$$

where $\hat{\varepsilon}_t$ is $y_t - X_t\beta$. (Note, the $\hat{\varepsilon}_t$ is not the least squares residual or the weighted least squares residual. It is the residual calculated using our (previous) random draw of β.) The likelihood as a function of $1/\sigma^2$ is therefore proportional to a Gamma$(\frac{T}{2}+1, \sum w_t \hat{\varepsilon}_t^2/(2\nu))$ pdf. (This implies

$$E[1/\sigma^2] = \frac{T+2}{\sum w_t \hat{\varepsilon}_t^2},$$

which makes sense.

Each individual w_t enters the likelihood in a separate factor, proportional to (with other parameters fixed)

$$\left(w_t^{\frac{\nu}{2}-1}\right) e^{-\frac{1}{2} w_t (1+\frac{\hat{\varepsilon}_t^2}{\sigma^2})}. \tag{5}$$

This is proportional to a Gamma$(\nu/2, \frac{1}{2}(\hat{\varepsilon}_t^2/\sigma^2+1))$ pdf, which again we know how to sample from.

As a function of ν, the likelihood is not in a handy form. One could take a "Metropolis-within-Gibbs" step to sample from it, if one used the $1/\nu^2$ prior, but for this exercise we will simply hold ν fixed at various values and see how inference changes. In order to compare marginal likelihood for different vaues of ν, we have to apply a modified harmonic mean or bridge sampling method to the sampled likelihood values, which you were not asked to do on the exercise.

So if you want to write your own code to implement the sampler you need for the exercise, it should be a loop with three stages

(i) Estimate β by weighted least squares and calculate the covariance matrix of β. Use these results to draw from the conditional normal distribution of β.
(ii) Form the residual vector from the regression and use it to make a draw from the inverse-gamma distribution of σ^2 (the pdf (4) above).
(iii) Draw a new $\{w_t\}$ sequence using the gamma distributions for the individual w_t's (the pdf's (5) above).

The `tshock()` function on the web site actually does these steps in the order (iii, i, ii), but of course the order doesn't matter. Also `tshock()` works with weights that are scaled by σ^2, instead of the w_t's in these notes.

A version of `tshock()` that includes comments blocking out these sections of the loop and explaining the use of the qr component of R's regression output is now on the web site. This version also generates log likelihood marginalized over the weight values instead of the log likelihood at particular weight vectors, which was what the previous version produced.