Boston University Questrom School of Business
MF840 – Spring 2019

Eric Jacquier

Problem Set 1
Due Thursday February 7th in class
Solutions

Problems turned in after the beginning of student section have a notch deduction.

Teams "*across sections*" turn in their homework at the beginning the morning class

Problems turned in after class get a zero.

- Do the Problem Set in groups of two
- Turn in one paper copy in class with two names, no electronic submission accepted.
- **To get a check, you need to answer all the questions.**
- **All discussion and theoretical questions must be hand written with a pen to count**.

Problem 1: *MLE and the Δ method*

Consider the regression of Y on X, T observations and k variables including the intercept, as in the notes. $Y = X\beta + \epsilon$. The errors are i.i.d. normal with mean zero and ***precision h****. The precision is simply $h = 1/\sigma^2$, the inverse of the variance. It is more and more often used in modern econometrics.*

a) Write the Log likelihood of the parameters (β , h) given the data (Y, X). To be clear, there can not be any σ or σ^2 in this density, only h! Then write the two first-order conditions with respect to β and h. Find the MLE estimates of β and h.

The log-likelihood is: $\text{Log}\ell = (T/2)\,\text{Log}\,h - h\,(Y-X\beta)'(Y-X\beta)/2$

$$\partial\,\text{Log}\ell/\partial h = T/(2h) - (Y-X\beta)'(Y-X\beta)/2 = 0 \qquad [1]$$

$$\hat{h}_{MLE} = \frac{T}{\hat{\epsilon}'\hat{\epsilon}}$$

$$\partial\,\text{Log}\ell/\partial\beta = -\,h\,(-2X'Y+2X'X\beta)/2 = 0, \text{ Of course no change or } \hat{\beta}_{MLE} \qquad [2]$$

b) Compute the 2nd order derivatives, then take the expectation, then compute the asymptotic covariance matrix of $\hat{\beta}_{MLE}$, $\hat{h}_{MLE}$. What is variance of $\hat{h}_{MLE}$?

$$\frac{\partial^2 \text{Log}\ell(\beta,v)}{\partial\beta^2} = -\,h\,X'X \qquad E\frac{\partial^2 \text{Log}\ell(\beta,v)}{\partial\beta\,\partial h} = E[-\ (-2X'Y + 2X'X\beta)/2] = 0$$

$$E\frac{\partial^2 \text{Log}\ell(\beta,v)}{\partial h^2} = E\frac{\partial}{\partial h}\left[\frac{T}{2h} - \frac{1}{2}(Y - X\beta)'(Y - X\beta)\right] = \frac{-T}{2h^2}$$

Now take the minus sign and invert, The off diagonal blocks are zero again. For the variance of $\hat{h}_{MLE}$, we get

$$V(\hat{h}_{MLE}) = \frac{2h^2}{T} \qquad V(\hat{\beta}_{MLE}) = \frac{1}{h}\left(X'X\right)^{-1}$$

c) Now use the Δ method to compute the asymptotic variance of the estimator of variance $\hat{v}_{MLE}$, where $v = 1/h$ is the variance of the regression noise.

$\partial v/\partial h = -1/h^2$ $\quad\quad V(\hat{v}_{MLE}) = -1/h^2 * \frac{2h^2}{T} * -1/h^2 = \frac{\mathbf{2}}{\mathbf{Th^2}} = \frac{\mathbf{2\sigma^4}}{\mathbf{T}}$

d) Now use the Δ method to compute the asymptotic variance of the estimator of $\boldsymbol{\sigma} = \sqrt{\boldsymbol{v}}$*, the* ***standard deviation*** *of the regression noise.*

$\partial\sigma/\partial v = v^{-0.5}/2$ $\quad\quad V(\hat{\sigma}_{MLE}) = v^{-0.5}/2 * \frac{2v^2}{T} * v^{-0.5}/2 = \frac{\mathbf{v}}{\mathbf{2T}} = \frac{\boldsymbol{\sigma^2}}{\mathbf{2T}}$

Problem 2: MLE estimation with heteroskedasticity

Stock returns are independently distributed: $r_t \sim N(\mu, \sigma_t)$. You think that the standard deviation σ_t is best approximated as a simple function of an observable strictly positive variable x_t: $\sigma_t^2 = \sigma^2 x_t$.
a) You collect T independent returns r_t's and corresponding x_t's Write the log-likelihood of the parameters given the data: $l(\mu, \sigma \mid R, X)$.

$$Log\ \ell \propto -T\ \log \sigma\sqrt{x_t}\ - \frac{1}{2\sigma^2}\sum \frac{(r_t - \mu)^2}{x_t} \propto -T\ \log \sigma\ - \frac{1}{2\sigma^2}\sum \frac{(r_t - \mu)^2}{x_t}$$

We don't need the x_ts in the log for optimization.

b) Maximize it to compute $\hat{\mu}_{MLE}$, and then $\hat{\sigma}_{MLE}$. Show your proof and result

$\partial\ \mathrm{Log}\ell/\partial\mu = 0: \quad -\frac{-2}{2\sigma^2}\sum\frac{(r_t-\mu)}{x_t} = 0$ $\quad\quad \hat{\boldsymbol{\mu}}_{\boldsymbol{MLE}} = \frac{\sum r_t/x_t}{\sum 1/x_t}$

$\partial\ \mathrm{Log}\ell/\partial\boldsymbol{\sigma} = \frac{-T}{\sigma} + \frac{1}{\sigma^3}\sum\frac{(r_t-\mu)^2}{x_t} = 0$ $\quad\quad \hat{\boldsymbol{v}}_{\boldsymbol{MLE}} = \frac{\mathbf{1}}{\boldsymbol{T}}\sum\frac{(r_t-\mu)^2}{x_t}$

c) Explain in a few words how this $\hat{\mu}_{MLE}$ is like a weighted least squares estimator. What are the weights? What observations are down- or over- weighted in the estimator, relative to the basic sample average?

Observations are weighted by their relative "precision" 1/x instead of equal weighted by 1/T. More precise observations are upweighted relative to 1/T, more noisy observation are downweighted.
The weighted sum is divided by the sum of the weights $\Sigma\, 1/x_t$ for the weights to sum to 1.

d) Use the Cramer-Rao lower bound to compute the asymptotic variance of $\hat{\mu}_{MLE}$.

Again, the cross derivative is zero: We can just compute each second derivative and change the sign. We have

$$\partial_2 \log\ell/\partial\mu^2 = \frac{1}{\sigma^2}\sum\frac{-1}{x_t}$$

$$V(\hat{\mu}_{MLE}) = \sigma^2 \Big/ \sum \frac{1}{x_t} = \frac{\sigma^2}{T} \frac{1}{\frac{1}{T}\sum \frac{1}{x_t}} = \frac{\sigma^2}{T} H(x_t)$$ where H is the harmonic mean

e) Compute the variance of the basic equal-weighted sample mean accounting for heteroskedasticity.

$$V(\hat{\mu}_{OLS}) = V\left(\sum \frac{r_t}{T}\right) = \frac{1}{T^2}\sigma^2 \sum x_t = \frac{\sigma^2}{T} A(x_t)$$

f)Compare the variances of the MLE and the equal-weighted naïve sample mean. You can use the results in the handout AM-GM-HM-inequalities.pdf to conclude.

The Harmonic mean of positive quantities is always smaller than their arithmetic mean

Problem 3: Feasible GLS

The file size-day-0918.csv contains the retuns on the 10 size-decile portfolios of the US stock market. For this exercise, you will use Decile 2. The goal is to estimate the portfolio beta and alpha. Use only the first 3 years of data for this. If you think you need more data, explain what it is and get it on KF's web site. Don't forget that KF's data are in %.

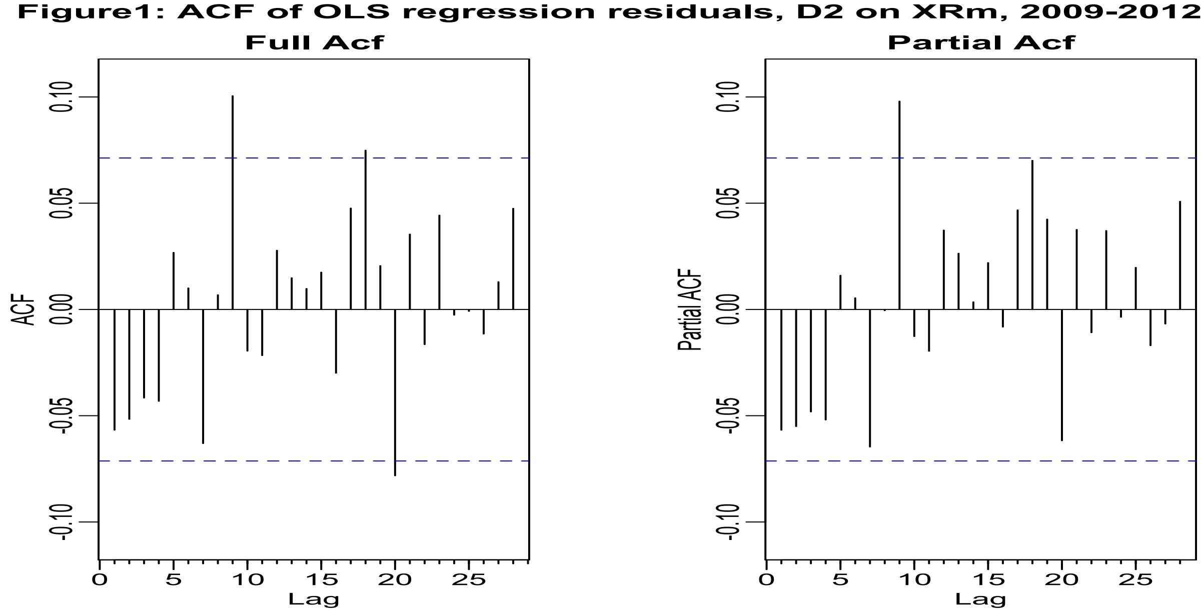
a) Estimate the portfolio's β and α, in a table put the estimates and OLS standard errors. In a second row, add (both types of) HAC standard errors. On the view of the OLS and HAC standard errors do you suspect a potential GLS situation?

> Comparing ordinary OLS standard errors and HC (heteroskedasticity) robust standard errors reveals a fair amount of heteroskedasticity. However it seems that incorporating autocorrelation does not affect the standard errors further.

b)

- *Plot the autocorrelations and partial autocorrelations of the residuals. Conclude as to the likeliest model (no need for a formal model search).*

Figure1: ACF of OLS regression residuals, D2 on XRm, 2009-2012



Probably an AR(0)!!

- *Report in a table the first four autocorrelations and their standard errors.*

The very approximate standard error is $1/\sqrt{T}$ = 0.036. Note the +/- 1.96 * 0.036 band on the plots. See MF793 notes. The first 4 ACFs are:

Lag 1	Lag 2	Lag 3	Lag 4
-0.06	0.05	-0.04	-0.04

- *Do a feasible GLS allowing for autocorrelation. Use an MA(3). That is, use directly the autocorrelations you just estimated to compute a TxT correlation matrix C.*

GLS-3ACF in the table uses the first three ACF as given without fitting an MA(3). We take the first three ACF above as our estimates and to built a correlation matrix with only 3 lags, longer lags all set to zero.

Tip: To get the covariance matrix, you also need to estimate the standard deviation of the noise and put it in a diagonal matrix D, then you can use the relation: $\Omega = D\,C\,D$.
Tip: You can fill up C in one line of code with the command "toeplitz". What is a Toeplitz matrix??

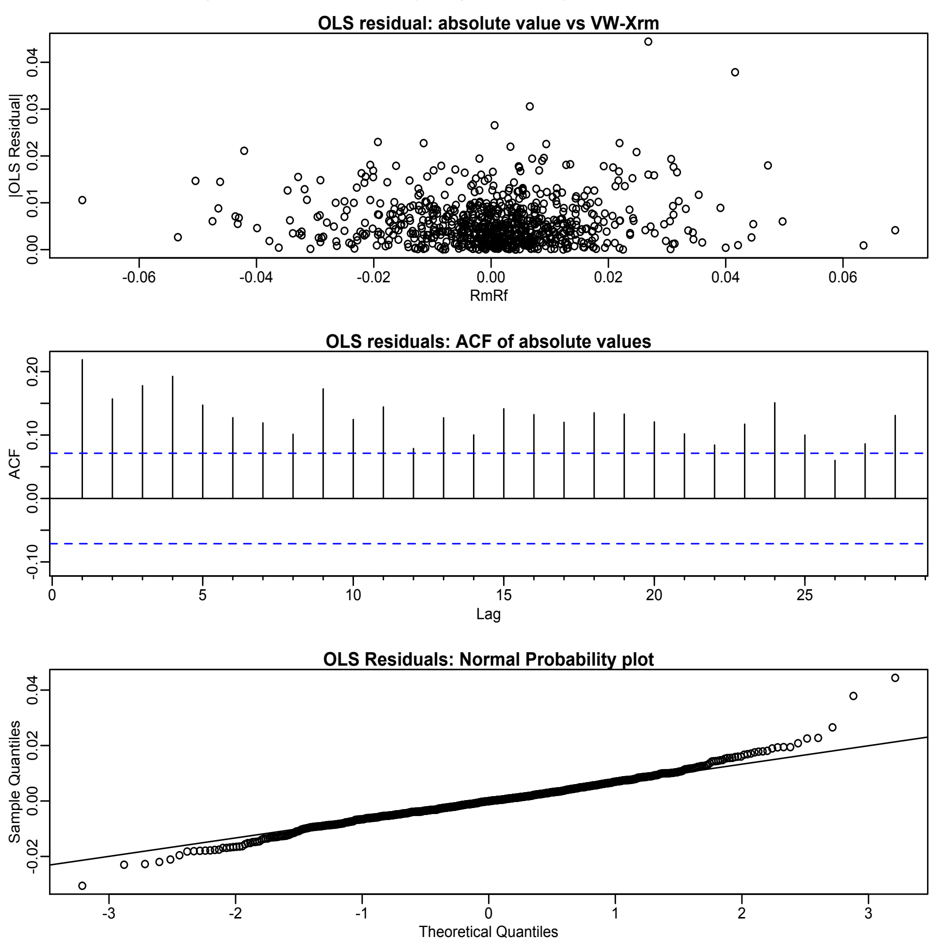
```
xx <-c( values of the first row of the matrix )
cc <-toeplitz(xx)  # Voila! Check that it work, read the manual
```

- *Add the β estimates and standard errors to your OLS results table*

c)

- *For your diagnostic of heteroskedasticity, do these two things: Plot absolute values of residuals vs X_t. Conclusion? For a diagnostic of potential GARCH effects, plot the acf and pacf of the absolute values of the residuals. Conclusion?*

Figure 2: Heteroskedasticity Diagnostic, Daily D2 on XRm, 2009-12



We do not detect an obvious relationship between variance and Xrm from plot a).

Plot b) shows a very strong persistence of absolute values which is a sign of predictable time varying variance.

Plot c) confirms what we expected, the strong non-normality of daily residual returns.

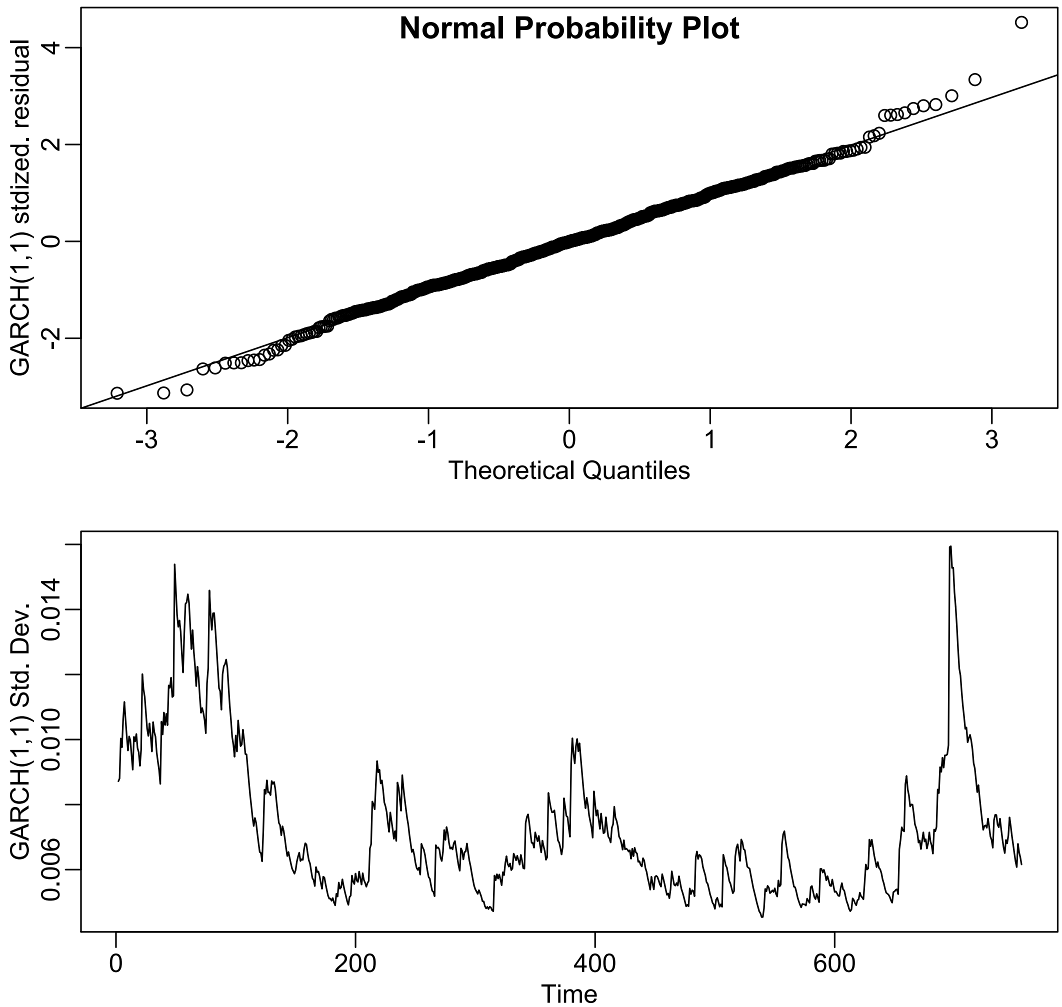
- *Estimate a GARCH(1,1) on the residuals of the OLS regression. Report the parameter estimates. Plot the time varying standard deviations of the residuals.*

Estimates are from command garch

	α_0	α_1	β_1
Estimate	1.24e-06	0.084	0.893
Standard Error		0.017	0.241

Even after regressions, residuals of daily returns also exhibit strong GARCH effects

Figure 3: GARCH(1,1) Diagnostics, Residuals of Daily D2 on XRm, 2009-12



- *Use these $\{\sigma_t\}$s to construct the Ω matrix and a feasible GLS with GARCH(1,1) errors.*
- *Add the β estimates and standard errors to the OLS results table.*

Table 1: Estimats of β				
	α	$s(\alpha)$	β	$s(\beta)$
OLS	*7.58e-05*	*0.000280*	***1.456***	*0.0188*
HC std. errors		*0.000279*		*0.0242*
HAC std. errors		*0.000266*		*0.0245*
GLS - 3ACF	*7.90e-05*	*0.00023*	***1.451***	*0.0188*
GLS – Garch(1,1)	*-7.07e-05*	*0.00024*	***1.408***	*0.0187*
GLS – Iter	*-5.20e-05*	*0.00031*	***1.427***	*0.0282*

Tip: Do not worry if you get a message of "failed convergence". Recall that the GARCH(1,1) can't estimate a variance for the first observation since it is an ARMA(1,1) in the squares. For the first observation, use the unconditional variance instead. That allows you to have a TxT matrix. Since GARCH gies you standard deviation estimates, you directly have the diagonal TxT covariance matrix.

d) Starting from c). Implement a simple loop to iterate your GLS. Save the estimate at each iteration. Stop the iteration when the % change in the estimates is less than 0.5%. Is the final estimate very different from the initial one?

It converges in 3 iterations.

e) Explain but do not do it, how you would implement a feasible GLS accounting for both autocorrelation and heteroskedasticity.

One practical idea is to do AR(2) and Garch(1,1) in sequence. You can alternate until it stabilizes. It's not theoretically the best thing to do because The estimation of one model may be biased by the presence of the other model.

Fortunately, we can write the likelihood of a ARMA with GARCH errors. So it is what should be done.