Boston University Questrom School of Business
MF 793 – Fall 2018

Eric Jacquier
Problem Set 5
Due Monday November 19th
The practical problem has been added.

- Problems turned in after the beginning of class have a notch deduction. Problems turned in after class are not graded.
- Do the Problem Set in groups of two
- Turn in a paper copy in your section of the class, no email submission accepted.
- To get a check, you need to answer all the questions including the discussion questions, and have your R code as a printed appendix at the end of the report.
- To get credit, everything which is not a plot, table, or R output **must be hand-written**.
- You can **not** get a check plus if you put irrelevant digits in your Tables. Think before you print.

A lot of this problem is geared at understanding the foundations in Lecture Notes 9 and 10.

Problem 1

LN 9 page 5, computes the ***true*** *optimal β when using with the linear model $y = \beta x + \epsilon$ to approximate an unknown CEF $E(y|x)$. Generalize the result on page 5 to the model*

$$y = \alpha + \beta x + \varepsilon \qquad [1]$$

That is find the true α and β, that minimize $E(\varepsilon^2|X)$.

$$y = \alpha + \beta x + \varepsilon \qquad E(\varepsilon) = E(x\varepsilon) = 0$$

Note: $E(y) = \alpha + \beta E(x) + 0$

$$\text{Min } E(\varepsilon^2) \iff \text{Min } E\left[y - E(y) - [x - E(x)]\beta\right]^2$$

$$\iff \text{Min } V(y) + \beta^2 V(x) - 2\,Cov(x,y)\beta$$

$$\frac{\partial \cdot}{\partial \beta} = 0 = 2\beta V(x) - 2\,Cov(x,y) \qquad \boxed{\beta = \frac{Cov(x,y)}{var(x)}}$$

Problem 2

Make sure you understand basic vector derivatives which arise in econometrics and multivariate optimization. Specially, you need to understand the "normal equation" on Lecture Note 10, page 9.

a) In the Sum of Squares, top of P. 9, what is the dimension of U=X'Y ? Set k=2 and consider a column vector U=(u_1, u_2)' and the scalar quantity q=β'U. Write out each component of the 2x1 vector $\partial q / \partial \boldsymbol{\beta} \equiv (\partial q/\partial\beta_1, \partial q/\partial\beta_2)$ *. Compare these two components to U.*

X'Y: X' is K x T, Y is T x 1, the product is K x 1. Let us just call it U and look at q = β'U.

$$\partial q/\partial\beta_1 = \partial(\beta_1 u_1+\beta_2 u_2)/\partial\beta_1 = u_1 \qquad \partial q/\partial\beta_2 = \partial(\beta_1 u_1+\beta_2 u_2)/\partial\beta_2 = u_2$$

The derivative of r with respect to β: $\partial q/\partial\beta$, is just the vector U=(u_1, u_2)'.

We can indeed write that: d(**β'U**)/d**β** = **U**.

b) Now let's study the derivative of a quadratic form, the second part of the sum of squares. Take a column vector β=(β_1, β_2)', and a matrix

$$Q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}.$$

*Write out (compute) the quadratic form in **β**, the scalar quantity H = **β**'Q**β**, as a function of the components of Q and **β**.*

$$H = (\beta_1 , \beta_2)\,(q_{11}\,\beta_1 + q_{12}\,\beta_2 , q_{21}\,\beta_1 + q_{22}\,\beta_2)' \quad = q_{11}\,\beta_1^2 + (q_{12} + q_{21})\,\beta_2\,\beta_1 + q_{22}\,\beta_2^2$$

c) Use the result in b) to write each component of the vector $(\partial H/\partial\beta_1, \partial H/\partial\beta_2) \equiv \partial H/\partial\boldsymbol{\beta}$*. Second, write each component of the 2x1 vector: 2Q**β**.*
*Compare the two computations: When is the derivative of H with respect to β equal to 2Q**β**?*

$$\partial H/\partial\beta_1 = 2\,q_{11}\,\beta_1 + (q_{12} + q_{21})\,\beta_2 \qquad \partial H/\partial\beta_2 = 2\,q_{22}\,\beta_2 + (q_{12} + q_{21})\,\beta_1$$

$$2Q\boldsymbol{\beta} = 2\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 2\begin{pmatrix} q_{11}\beta_1 + q_{12}\beta_2 \\ q_{21}\beta_1 + q_{22}\beta_2 \end{pmatrix}$$

The derivative of H with respect to β equals 2Qβ **when the matrix Q in the quadratic form is symmetric .... like a covariance matrix !**

<u>Problem 3</u>: OLS estimator, one variable plus one intercept

What is inside (X'X)[-1] ? You need to have done it once in the 2x2 or 3x3 case! Take the OLS estimator $\hat{\beta}_{OLS}$=

(X'X)$^{-1}$*X'Y and its variance* $\mathbf{V(\hat{\beta}_{OLS})}$ = σ^2*(X'X)*$^{-1}$*. For simple regression with intercept and only one X variable,* $\boldsymbol{\beta} = (\alpha, \beta)$*.*

a) *In the variance formula, write each component of the 2x2 matrix (X'X)*$^{-1}$*, using only the sample statistics:* $\bar{X}, \bar{Y}, \widehat{\sigma_{XY}}, \widehat{\sigma_X}, \widehat{\sigma_Y}$*, and the sample size T.*

b) *Do the same for* $\hat{\beta}_{OLS}$*: write its 2 components only as functions of the sample statistics and T.*

$$X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_T \end{pmatrix} \qquad X'X = \begin{pmatrix} T & T\bar{x} \\ T\bar{x} & T(\hat{\sigma}_x^2 + \bar{x}^2) \end{pmatrix}$$

because $\hat{\sigma}_x^2 = \frac{1}{T}\sum x_t^2 - \bar{x}^2$

$$\sigma^2 (X'X)^{-1} = \frac{\sigma^2 \, T}{T^2 \hat{\sigma}_x^2} \begin{pmatrix} \hat{\sigma}_x^2 + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_T \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix} = \begin{pmatrix} T\bar{Y} \\ \sum x_t Y_t \end{pmatrix}$$

$$\hat{\beta} = \frac{1}{\hat{\sigma}_x^2 T} \begin{pmatrix} T\bar{y}\hat{\sigma}_x^2 + T\bar{y}\bar{x}^2 - \bar{x}\sum x_t Y_t \\ -T\bar{x}\bar{y} + \sum x_t Y_t \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \bar{x}\left(\cancel{T}\bar{y}\bar{x} - \frac{1}{T}\sum x_t Y_t\right)/\hat{\sigma}_x^2 \\ \widehat{Cov(x,y)}/\widehat{\sigma_x^2} \end{pmatrix} = \begin{pmatrix} \bar{y} - \bar{x}\hat{\beta}_1 \\ \hat{\beta}_1 \end{pmatrix}$$

c) *Look at V(*$\hat{\beta}_{OLS}$*), what leads to a more precisely estimated* ***slope*** *coefficient?*

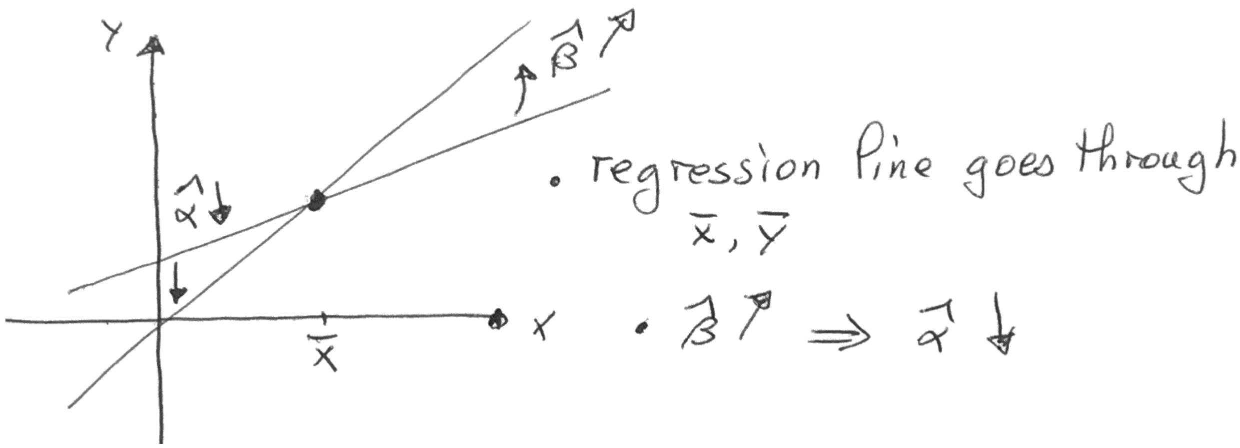
A large sample size T, a small noise variance σ^2 as expected. … But also:
A large sample variance of the X data leads to a more precisely estimated slope coefficient.

d) *Given the Cov(a,b), you just computed: What is likely to happen to the estimate of the intercept if we overestimate the slope? You can support your answer with a simple hand drawn picture of Y vs X: Remember that the regression line goes through* $(\bar{x}, \bar{y})$*, look at Cov(a,b) found in question a. There will be two cases according to the sign of* $\bar{x}$*.*

We see that: $\mathbf{Cov(\hat{\alpha}, \hat{\beta})}$ **is proportional to** $-\ \bar{\mathbf{x}}$**.**

If the sample mean of the x data is positive, then the covariance of the slope and intercept estimators is negative. This means that if one overestimates the slope, one will likely underestimate the intercept.

When $\bar{x}$ is negative, we have the opposite result. See the picture below with $\bar{x} > 0$



<u>Problem 4:</u> Estimating portfolio vs stock Market betas.
Consider N stocks which returns R_i have standard deviation σ_I, beta with the market portfolio β_i. The market portfolio has standard deviation σ_m. Consider a portfolio P of the N stocks with weights w_i. β is estimated by the usual single regression of the asset on the market return.
a) Using the obvious formula for β for example as you found in Problem 1

$$\beta_P = \sum_{i=1}^{N} w_i \beta_i \qquad [1]$$

b) So the β of a portfolio is …. a weighted average of the stock betas, the same way the portfolio return is a weighted average of the stock returns. [1] also applies to the estimators of β, so $b_P = \sum_{i=1}^{N} w_i b$. The estimators are random variables. We use this to get an insight on the variance of the estimator of a portfolio beta:

Consider an equal weighted portfolio of N stocks. The stocks are all equally correlated $\rho_{ij} = \rho$, $\forall$ i,j. For simplicity, assume that the estimators b_i all have the same variance V_b. It can be proven but just take the result as give that all pairs of estimators b_i ,b_j are equally correlated, call it ρ_b. Use the portfolio variance formula in Lecture Note 3 to prove that:

$$Var(b_p) = V_b \, [\, 1/N + (1 - 1/N)\, \rho_b \,] \qquad [2]$$

a) $\beta_p = \dfrac{Cov(\tilde{R}_P, \tilde{R}_M)}{\sigma_M^2} = \dfrac{Cov\left(\sum_{i=1}^{N} w_i \tilde{R}_i, \tilde{R}_M\right)}{\sigma_M^2}$

$$\beta_P = \frac{\sum w_i \, Cov(\tilde{R}_i, \tilde{R}_M)}{\sigma_M^2} = \sum w_i \beta_i$$

b) Simplifying assumption just makes calculus easier, it does not change the fundamental result.

$$Var(\hat{\beta}_P) = Var\left(\sum_i \frac{1}{N} \hat{\beta}_i\right)$$

$$= \frac{1}{N^2} \sum_i v(\hat{\beta}_i) + \sum_i \sum_{j \neq i} \frac{1}{N} \frac{1}{N} cov(\hat{\beta}_i \hat{\beta}_j)$$

$$= \frac{1}{N^2} N v(\hat{\beta}) + \frac{N(N-1)}{N^2} \rho_b v(\hat{\beta})$$

$$var(\hat{\beta}_P) = var(\hat{\beta}) \left[\frac{1}{N} + \left(1 - \frac{1}{N}\right) \rho_b \right]$$

c) Explain what [2] means for the precision of the estimate of a portfolio beta relative to the precision of the estimate of the beta of the "typical" stock in the portfolio.

Since ρ_b<1, portfolio betas are estimated more precisely than stock betas, and more precisely the larger they are.
How small can ρ_b get? It's all in these numbers from Gene Fama's first regressions, 20 portfolios of about 100 stocks each.

ESTIMATION PERIOD 1962-1966

b_{pm}	.514	.625	.665	.697	.791	.812	.843	.888	.916	.940	.941	.943	.976	1.062	1.070	1.216	1.291	1.316	1.365	1.486
$s(b_{pm})$	.042	.046	.050	.040	.032	.032	.034	.033	.029	.034	.032	.030	.038	.036	.036	.041	.032	.046	.052	.056
$\hat{\rho}(R_p, R_m)^2$	.72	.75	.75	.84	.91	.92	.91	.92	.94	.93	.94	.94	.92	.93	.94	.94	.96	.93	.92	.92
$s(R_p)$	.025	.030	.032	.032	.035	.036	.037	.039	.039	.041	.041	.041	.043	.046	.046	.053	.055	.057	.060	.065
$s(e_p)$	.013	.015	.016	.013	.010	.010	.011	.011	.009	.011	.010	.010	.012	.012	.012	.013	.010	.015	.017	.018
$\bar{s}_p(e_j)$	.053	.046	.046	.053	.057	.057	.059	.057	.055	.060	.053	.056	.064	.062	.067	.075	.068	.071	.076	.089
$s(e_p)/\bar{s}_p(e_j)$	.24	.32	.34	.24	.17	.17	.18	.19	.16	.18	.18	.17	.18	.19	.17	.17	.14	.21	.22	.20

SOURCE: Eugene F. Fama and James D. MacBeth, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* 71 (May-June 1973): 620-621.

<u>Problem 5:</u> Forecasting Error Variance

Starting from Result [1], page 6 of LN 11, prove the red result on page 7. Do NOT use the method suggested on page 7. Instead, just use your results in Problem 3 where X contains one intercept and explanatory variable. Rewrite the 2x1 vector $\mathbf{x_f}$ as $(1,x_f)'$.

$$V(e_f) = \sigma^2\left(1 + \underline{x}_f\,(X'X)^{-1}\,\underline{x}_f\right)$$

$$= \sigma^2\left(1 + \frac{(1\ \ x_f)}{T\sigma_x^2}\begin{pmatrix}\sigma_x^2+\bar{x}^2 - x_f\bar{x}\\ x_f - \bar{x}\end{pmatrix}\right)$$

$$= \sigma^2\left(1 + \frac{1}{T\sigma_x^2}\left[\sigma_x^2 + \bar{x}^2 + f_x^2 - 2x_f\bar{x}\right]\right)$$

$$= \sigma^2\left(1 + \frac{1}{T} + \frac{(x_f-\bar{x})^2}{T\sigma_x^2}\right)$$

Problem 6: Sampling Frequency and Portfolios

Get the 11 stock monthly and daily returns. Make sure to download the daily file again because I modified vwret which had a timing issue. Make the 11 stock portfolio. Now regress them all on the market return and fill Table 1

Table 1:								
	Monthly Regressions				Daily Regressions			
	α **x12**	t_α	β	s_β	α **x252**	t_α	β	s_β
Apple	0.14	1.73	0.96	0.18	0.14	1.88	0.92	0.03
Amazon	0.17	1.92	1.08	0.20	0.18	1.88	1.08	0.04
Biogen	0.15	1.50	0.99	0.23	0.15	1.53	1.02	0.04
Citygroup	-0.07	-1.05	1.74	0.16	-0.06	-0.79	1.66	0.03
GE	-0.08	-1.23	1.18	0.14	-0.07	-1.28	1.07	0.02
Nike	0.11	1.62	0.70	0.16	0.09	1.35	0.88	0.03
Pepsi	0.07	1.74	0.43	0.09	0.06	1.43	0.50	0.02
State Street	-0.03	-0.51	1.43	0.15	-0.02	-0.34	1.34	0.03
Toyota	0.00	-0.02	0.71	0.14	-0.02	-0.26	0.84	0.03
Valero	0.10	1.02	1.58	0.23	0.12	1.29	1.41	0.04
Verizon	0.07	1.24	0.44	0.14	0.05	1.08	0.58	0.02
Average	0.06	0.72	1.02	0.17	0.06	0.71	1.03	0.03
EW Portfolio	0.06	2.58	1.02	0.05	0.06	2.86	1.03	0.01

In Row Average, put the average of the statistics for the 11 stocks. Use the row Average to build intuition to answer the questions below.

Note that Table 1 here shows the Student-t, not the standard deviations.

a) Compare the 12 monthly and daily returns Betas.

- *Are the βs from both frequencies markedly different?*

  Monthly and Daily β estimates are not very different. There is no evidence of underestimation of daily βs

- *Are their standard errors markedly different?*

  Daily s_βs are 0.17/0.03 ≈ 6 times smaller than monthly s_βs. It's the order of magnitude expected for precision due to (dis)aggregation from monthly to daily data. ($\sqrt{21}$).

- *Which frequency seems best as an estimation strategy?*

  So, the higher frequency seems the best strategy for these liquid and synchroneous (with the market index) stocks.

- *For other stocks than these very large US stocks, what could be a problem with estimating βs from daily returns regressions?*

  In other situations, the problem we may worry about is either illiquidity or non-synchroneity such as returns in different time zones.

b) Compare the 12 monthly and daily returns intercepts. You can help your intuition by looking at the t-statistics outputs for the monthly vs daily returns regressions.

- *Are the estimates very different (after appropriate transformations)? Feel free to transform the α and s_α columns as you think makes more sense for the reader.*

    Once again the estimates are not systematically different across between monthly and daily returns regressions.

    The difference is due to the fact that αs are analogous to mean (excess) returns. After annualizing both, one sees that the monthly and daily frequencies give similar estimates.

- *Are the standard errors very different (after appropriate transformation)? Explain why the result differs from the β estimation.*

    The smaller standard errors of the daily regressions are due to the implicit daily horizon of the returns. You should be able to use the estimates and s_α to compute properly annualized confidence intervals for the monthly or annual α implied by the daily regression. Then you would see that there is no gain in precision over the monthly regression.

    As the homework mentioned, you get the same intuition by looking at the t-statistics from both regressions.

c) Monthly regression: portfolio vs stock regression precision

- *What is the average $\widehat{\beta}$ standard error for the 11 stock regression? What is the standard error for the EW portfolio regression $\widehat{\beta}$?*

    The average s_β is **0.17**, the portfolio s_β is **0.05**.

- Use equation [2] in Problem 4 to infer an average cross-correlation of the 11 $\widehat{\beta}$.

    You would save time by answering the question below which shows that with ρ=0, we would get about the ratio of 0.17/0.05. But taking the estimates as given, equation [2] yields:

    $(11*(0.507/0.165)^2-1)/(11-1)$ = **0.003**

- *If the correlation ρ_b was exactly zero, how smaller do you expect the standard deviation of the $\widehat{\beta}$ of an 11 stock portfolio to be than that of a typical stock.*

    With ρ = 0, we expect the portfolio s_β to be sqrt(1/11) = 0.3 that of a stocks s_β. So for a stock s_β = 0.165, we expect **0.050**

- *Get the 11 residual vectors (of the 11 stock regressions) in one data matrix. Compute their 11x11 correlation matrix, and compute the average cross-correlation. Make sure to remove the ones. What is it? Compare to your answer above.*

    You should have found **-0.001 !**

    For comparison, the average cross-correlation of the 11 stocks is **0.27**

    The market model regression explains about all the average cross-correlation of stock returns over 2010-2017

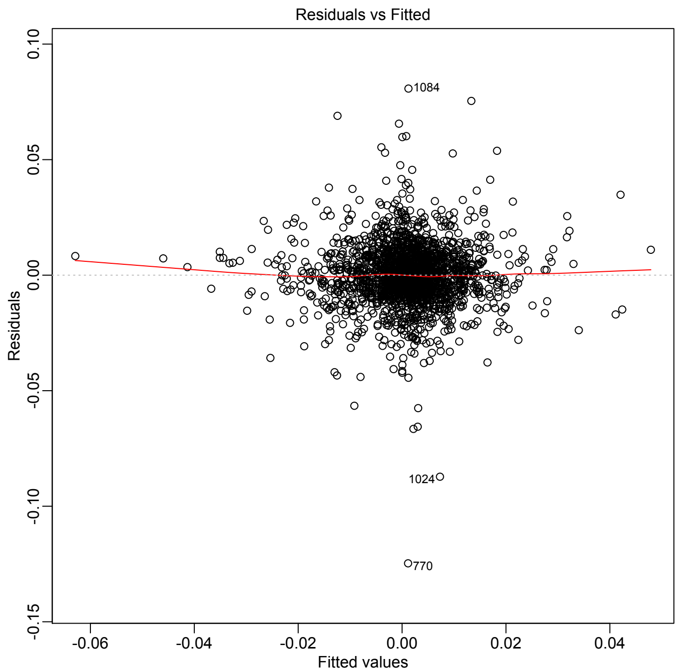
Problem 7: Typical Output Analysis

*Consider the AAPL regression on **daily** returns. For each diagnostic plot, indicate what it is supposed to detect, and if you find anything suspicious for the regression. You don't need to show a normal probability* *plot but check for yourself that it is very non normal.*

a) Standardized residuals vs the market return in Figure 1.

To be very clear, both lm$residuals and summary.lm$residuals give the same ordinary residuals, **non-standardized**.
For a "well-behaved" regression, the diagonal elements of P are all close to K/N. So the naively- and the properly-standardized residuals look very similar.

In R
rstandard(mymodel) gives you the properly standardized (with the hat matrix) residual.
rstudent(mymodel) uses the external noise σ.
Look at the R solution for more information



The plot here from *plot(mymodel)* uses the ordinary residuals.

Doing it yourself (vs X, not vs the fitted value) and using the standardized residuals:

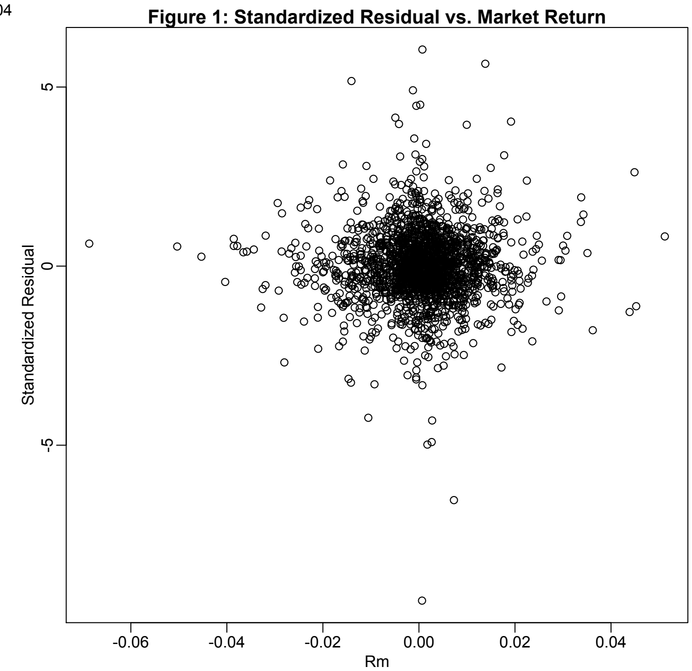
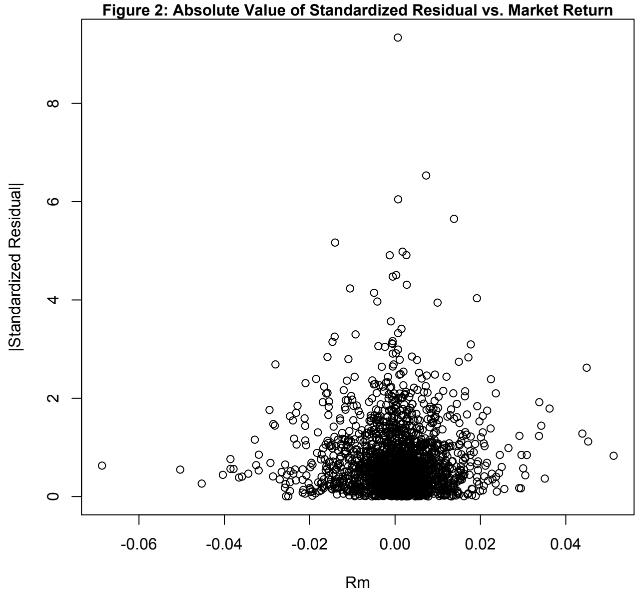


Figure 1: Standardized Residual vs. Market Return

These plots help detect a relationship between residuals and the X variables, typically a sign of an omitted variable correlated with an included variable, or of an incorrect functional form.

b) Absolute value of the standardized residuals vs the market return in Figure 2.

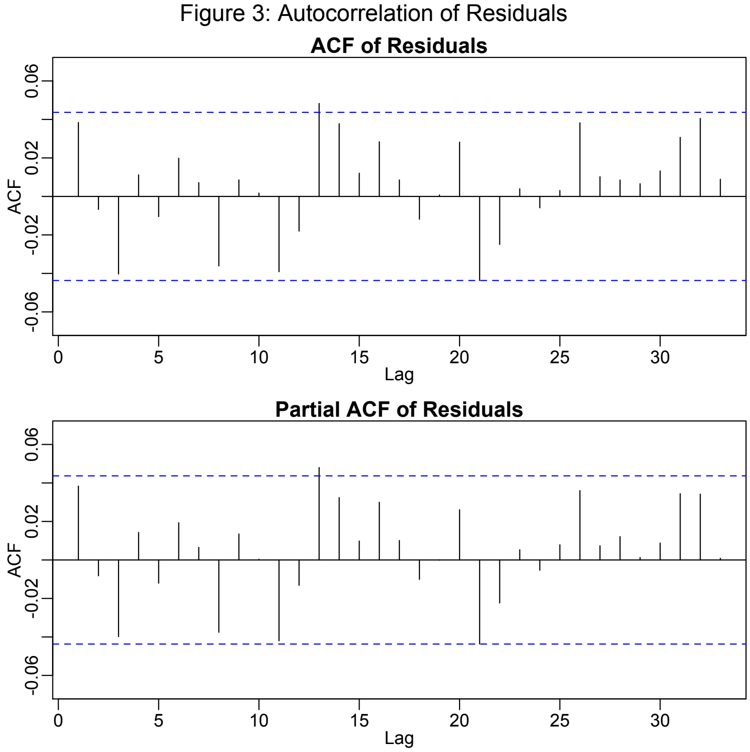
Figure 2: Absolute Value of Standardized Residual vs. Market Return



By concentrating on the absolute value, we can possibly detect a form of heteroskedasticity related to the X variables. Nothing really apparent here.

*d) Autocorrelation function of the standardized residuals in Figure 3. Do not use the standard **acf** command. Instead, use the library: **forecast**, and use the **Acf** command (with uppercase A)***.**

Figure 3: Autocorrelation of Residuals



There is no apparent autocorrelation in the residuals

e) Fill in the table and conclude whether heteroskedasticity is an issue. Load packages lmtest (to run coeftest and coefci) and sandwich (to run the adjusted covariance matrices). The (2,2) element of Cov is the variance of $\hat{\beta}$, the slope coefficient.
Relevant commands are: vcov(model) vcovHC(model) vcovHAC(model) coeftest(model)
coeftest(mm,vcov=vcovHC) etc… coefci(model,…) for a confidence interval

Table 2: Sandwich estimates of the slope standard error for the APPL daily regression

	$\hat{\beta}$	$s\hat{\beta}$	Cov[2,2]	5%	95%
OLS – iid	0.924	0.0317	1.004 E^{-03}	0.872	0.977
HC – White	0.924	0.0333	1.112 E^{-03}	0.870	0.979

There is no meaningful difference in the standard error of $\hat{\beta}$ between the OLS formula and the Heteroskedasticity consistent formula. However Don Andrew's correction does show a difference.

For Problem 6, you can loop around lm for the 12 regressions. You can also do them all with:
mymod <- lm(stkrets~rm) # creates the 12 regressions
modsum<-summary(mymod) # creates the summary object with all the good stuff in it.
coefficients(mymod) # also works to print the output with estimates and std.errs.

*But mymod is a **multivariate** linear model object, a list where each regression is an item in the list, then each regression is itself a list. Same for modsum. It makes it hard to retrieve in a vector for example, all the standard errors. To see the problem, do names(mymod), names(modsum).*
Some commands don't work. You can't use things like: confint(mymod)
What we want is not two steps down in the hierarchy of the list (of list)
*You can extract it with the **lapply** (ell apply!) and **sapply** commands. They are similar to **apply** but work to extract components of lists. In coefficients(mymod), you see in what order the output is. Now try this for example, you will see what it does:*

lapply(modsum, coefficients) is exactly the same as coefficients(modsum). But is is now a list with 12 items. This takes what we need from it:

```
sapply(lapply(modsum,coefficients),'[',c(2,4))
```

coefficients output for all regressions ‘[‘ says to go down one level c(2,4) takes elements 2 and 4