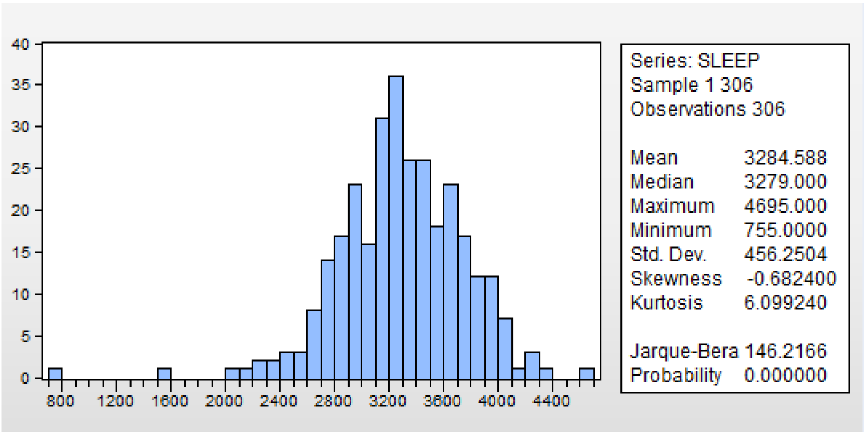
Question 2 (Final Exam, 2016 S1)

We would like to make a predictive model for amount of time that women sleep at night. We have randomly selected 306 women and asked them to record number of minutes slept every night for a week, the amount of hours that they engaged in paid work in that week, and also asked some of their personal characteristics such as age, education and number and age of their children. The variables that we will use in the analysis are:

$SLEEP$	minutes sleep at night, per week
WRK	minutes paid work, per week
$EDUC$	years of education
AGE	age in years
KID	=1 if any children under 3 years old, 0 otherwise

a. Based on preliminary analysis of the data, in particular based on the information in the following figure, we have decided to drop the observation that reports 755 minutes of sleep in the survey week. Explain why you agree or disagree with this decision. [2 marks]



We have estimated the following regression (the standard errors are reported below the parameter estimates)

$$\begin{aligned}
\widehat{SLEEP} &= \underset{(366.51)}{4523.48} - \underset{(0.03)}{0.13}\ WRK - \underset{(17.62)}{43.00}\ AGE + \underset{(0.21)}{0.50}\ AGE^2 - \underset{(9.00)}{12.50}\ EDUC - \underset{(88.12)}{144.54}\ KID \\
n &= 305,\ R^2 = 0.104 \qquad (1)
\end{aligned}$$

b. Test the null hypothesis that all else constant, having a child under the age of 3 has no effect on a mother's sleep against an alternative that makes sense in this context. Perform the test at the 5% level of significance and decide if you would or would not drop KID from the regression. [3 marks]

c. Compute a 95% confidence interval for the difference between the mean sleep time per week for two women with the same age and no young children, who work the exact same hours in a week, but one has 12 years of education and the other 16 years of education. [3 marks]

d. Explain the insights that the regression results provide for the effect of age on sleep, all else equal. In particular, all else equal, at what age women are predicted to sleep the least on average according to this estimated equation? [3 marks]

Our research assistant has estimated a series of regressions which are reported below (after part (f)). In these regressions $YHAT$ and $UHAT$ refer to the predicted values and residuals of equation (1).

e. Using the appropriate equation or equations, test that $EDUC$ and KID are jointly insignificant for predicting $SLEEP$ at the 5% level of significance. [3 marks]

f. Using the appropriate equation or equations, test for heteroskedasticity in the errors of equation (1). Perform the test at the 5% level of significance, and based on your conclusion, explain if the OLS estimators in (1) are unbiased and if the test results and confidence intervals computed in previous parts are reliable. [3 marks]

$$\begin{aligned} \widehat{SLEEP} &= \underset{(50.64)}{3519.43} - \underset{(0.03)}{0.13}\ WRK \\ n &= 305,\ R^2 = 0.078 \end{aligned} \tag{2}$$

$$\begin{aligned} \widehat{SLEEP} &= \underset{(333.60)}{4206.17} - \underset{(0.03)}{0.13}\ WRK - \underset{(17.46)}{37.64}\ AGE + \underset{(0.21)}{0.47}\ AGE^2 \\ n &= 305,\ R^2 = 0.092 \end{aligned} \tag{3}$$

$$\begin{aligned} \widehat{UHAT} &= \underset{(9048.98)}{-8462.416} + \underset{(5.45)}{5.10}\ YHAT - \underset{(0.001)}{0.001}\ YHAT^2 \\ n &= 305,\ R^2 = 0.003 \end{aligned} \tag{4}$$

$$\begin{aligned} \widehat{UHAT^2} &= \underset{(6275058)}{-2590915} + \underset{(3779)}{1696}\ YHAT - \underset{(0.568)}{0.260}\ YHAT^2 \\ n &= 305,\ R^2 = 0.001 \end{aligned} \tag{5}$$

<u>Question 2</u> (Final Exam, 2016 S2)

2.a. In the multiple regression model

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times(k+1)}{\mathbf{X}} \underset{(k+1)\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{u}},$$

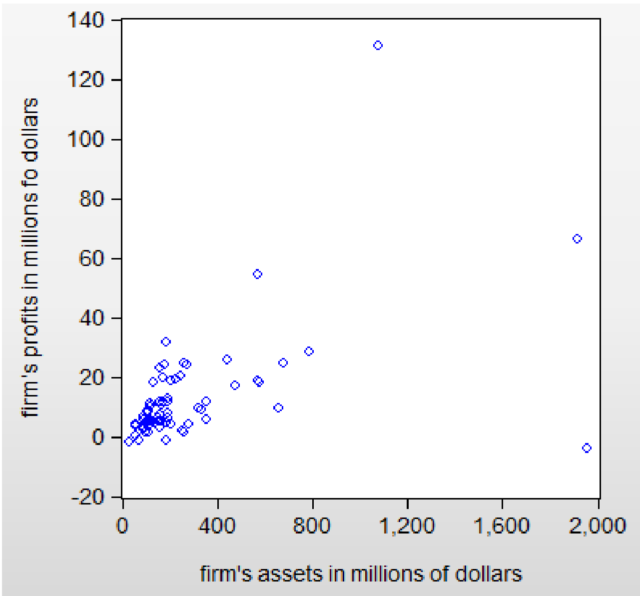
state the assumptions necessary for the OLS estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ to be an unbiased estimator of $\boldsymbol{\beta}$. Provide a proof of unbiasedness of $\widehat{\boldsymbol{\beta}}$ and indicate where each of these assumptions is used in your proof.

(5 marks)

2.b. We want to know if the separation of corporate management from corporate ownership affects a firm's performance after controlling for its assets. We have data on profits and assets for a randomly chosen cross section of 69 firms. 37 of these firms are managed by their owners and the other 32 are managed by professional managers who are not the owners of the firm. The summary statistics for profits and sales are provided in the table below.

Sample statistic	$PROFITS$ (million \$)	$ASSETS$ (million \$)
Mean	12.8	277.2
Median	7.4	168.4
Minimum	-3.8	30.3
Maximum	131.0	1953.2
Standard deviation	18.6	345.8

The scatter plot of profits versus assets is shown in the figure below



2.b.i) We run a simple regression of $PROFITS$ on a constant and $ASSETS$. What does the scatter plot suggest about (a) the sign of the slope coefficient? (b) the conditional variance of the errors? What problems would conditional heteroskedasticity cause for inference in this regression and is it likely that a log-log formulation would solve the heteroskedasticity problem in this application?

(5 marks)

2.b.ii) We have defined a dummy variable MNO which is equal to 1 if the firm is managed by a manager who is not the owner of the firm, and is equal to 0 otherwise. In order to get the best linear unbiased estimators of all parameters in

$$PROFITS = \beta_0 + \beta_1 ASSETS + \beta_2 MNO + u \tag{6}$$

and comment on whether the separation of management from ownership affects a firm's performance, we have estimated the following equation:

$$(W \times \widehat{PROFITS}) = \underset{(1.70)}{2.49} W + \underset{(0.01)}{0.04} \; (W \times ASSETS) - \underset{(1.88)}{2.03} \; (W \times MNO)$$

where $W = \frac{1}{\sqrt{ASSETS}}$. Under what assumption about the conditional variance of u in (6) would this estimated equation provide the best linear unbiased estimator of the parameters of β_0, β_1 and β_2? Given that assumption, explain why multiplying equation (6) by W produces a model that satisfies all requirements of the Gauss-Markov Theorem.

(5 marks)

2.b.iii) We want to make sure that the form of management affects neither the intercept nor the slope of the conditional expectation of $PROFITS$ as a function of $ASSETS$. For that reason, we would like to test that $\beta_2 = \beta_3 = 0$ in

$$PROFITS = \beta_0 + \beta_1 ASSETS + \beta_2 MNO + \beta_3 (MNO \times ASSETS) + u,$$

against the alternative that at least one of them is not zero. From preliminary analysis we have concluded that errors are heteroskedastic and weighting all variables by $W = \frac{1}{\sqrt{ASSETS}}$ solves the heteroskedasticity problem. Our research assistant has provided us with the following estimation results:

$$\begin{aligned} \widehat{PROFITS} &= 5.19 + 0.03\ ASSETS \\ \hat{\sigma} &= 16.11,\ SSR = 17392 \end{aligned} \tag{7}$$

$$\begin{aligned} (W \times \widehat{PROFITS}) &= 1.40W + 0.04\ (W \times ASSETS) \\ \hat{\sigma} &= 0.66,\ SSR = 27.95 \end{aligned} \tag{8}$$

$$\begin{aligned} \widehat{PROFITS} &= 1.56 + 0.05\ ASSETS + 8.23MNO \\ &\quad -0.05(MNO \times ASSETS), \\ \hat{\sigma} &= 13.17,\ SSR = 11272 \end{aligned} \tag{9}$$

$$\begin{aligned} (W \times \widehat{PROFITS}) &= 0.12W + 0.06\ (W \times ASSETS) + 2.50\,(W \times MNO) \\ &\quad -0.03\,(W \times MNO \times ASSETS)\,, \\ \hat{\sigma} &= 0.62,\ SSR = 25.32 \end{aligned} \tag{10}$$

Use the appropriate information to test $\beta_2 = \beta_3 = 0$ at the 1% level of significance. Remember to state the null, the alternative, the test statistic and its distribution under the null and the rejection rule, and state your conclusion about the effect of separation of management from ownership on a firm's performance.

(5 marks)

Question 3 **(Final Exam S2, 2017)**

3.a. We wold like to estimate the demand for beef. We have a time series sample of 150 seasonally adjusted quarterly observations on

$$\begin{aligned} Q_{beef} &= \text{quantity of beef demanded in kg} \\ P_{beef} &= \text{price of beef in dollars per kg} \\ P_{chicken} &= \text{price of chicken in dollars per kg} \\ INC &= \text{disposable income in dollars.} \end{aligned}$$

We consider the demand model for beef written as:

$$Q_{beef,t} = \beta_0 + \beta_1 P_{beef,t} + \beta_2 P_{chicken,t} + \beta_3 INC_t + \beta_4 t + u_t, \tag{11}$$

where t is a time trend such that $t = 1, 2, \ldots, 150$. If (11) is dynamically well specified, what type of process would you expect the error term (u_t) to follow? Provide the properties of this process.

(2 marks)

3.b. The researcher estimated (11) by OLS and obtained the corresponding residuals, denoted by $\hat{u}_t$. He plotted the sample autocorrelation and partial autocorrelation functions for $\hat{u}_t$, which are shown in Figure 1 below:

Figure 1

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.556	0.556	62.684	0.000
		2	0.436	0.184	101.43	0.000
		3	0.317	0.026	122.04	0.000
		4	0.183	-0.07...	128.98	0.000
		5	0.124	-0.00...	132.18	0.000
		6	0.112	0.054	134.79	0.000
		7	0.098	0.033	136.81	0.000
		8	0.051	-0.04...	137.35	0.000
		9	0.070	0.039	138.38	0.000
		1...	0.063	0.020	139.22	0.000

(i) What does the information in Figure 1 suggest with regard to the behaviour of the error term in (11)? Briefly explain.

(2 marks)

(ii) Set up a Breusch-Godfrey test that can be used for testing no serial correlation in errors against the alternative of serial correlation of order 2. Clearly state the steps involved, the null and alternative hypotheses of the test, the statistic(s) of interest and corresponding distribution(s).

(3 marks)

(iii) The R^2 obtained from the regression estimated in 3.b.(ii) with $\hat{u}_t$ as dependent variable was equal to 0.658. What conclusion would you draw with respect to the behaviour of the error term? Briefly explain.

(2 marks)

(iv) What are the implications of the results drawn in 3.b.(iii) with regard to linear regression modelling? What standard errors for the OLS estimates would you recommend using in this instance? Briefly explain.

(3 marks)

3.c. Our economist friend tells us that economic theory suggests the following nonlinear demand function for beef at time t:

$$Q_{beef,t} = e^{\gamma_0+\delta t+v_t} P_{beef,t}^{\gamma_1} P_{chicken,t}^{\gamma_2} INC_t^{\gamma_3}, \tag{12}$$

where variables Q_{beef}, P_{beef}, $P_{chicken}$, INC and t are defined above and v_t is a random error term. He tells us that as a result, the linear regression technique would be incapable of providing estimates of the parameters of this demand function, namely γ_1, γ_2 and γ_3, and we need more sophisticated nonlinear estimation techniques.

(i) Explain to him that he is incorrect by recommending an appropriate transformation of (12).

(2 marks)

(ii) Using the transformed model, or otherwise, interpret the parameter γ_1.

(1 mark)