OLS-parameter estimation

Bernt Arne Ødegaard

24 November 2021

Contents

1 The geometry of Least Squares.

1.1 Introduction

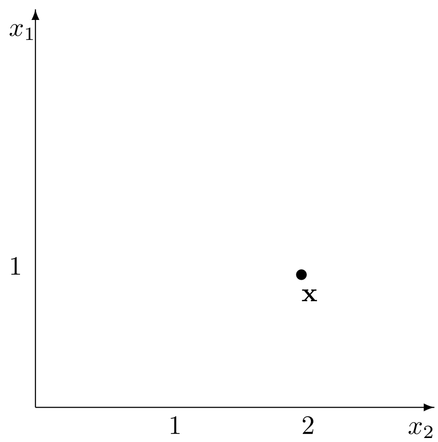
We will start by looking at the intuition behind the most common estimation method, the method of Least Squares. Since a lot of the estimation methods we use are based on least squares, the payoff from having some geometric intuition is large. We will thus spend some time on geometric and other intuition, without discussion the probabilistic background. This material only concerns how to find point estimates. To evaluate the significance of any point estimates we must work harder, using concepts from probability theory. But for the mechanics of finding an estimate it is not necessary to do any probability theory.

1.2 OLS estimation, simplest possible case

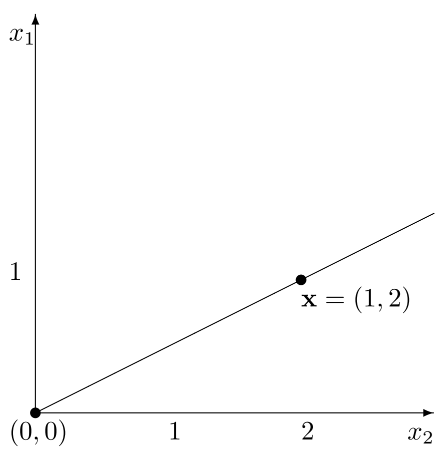
Consider the vector

$$\mathbf{x} = (1, 2)$$

In a plot, it is a point in a two-dimensional space



By taking linear combinations, which in this case is to multiply $\mathbf{x}$ with a constant β, we can generate a *line* in this figure that passes trough $\mathbf{x}$ and the origin $(0, 0)$.



This is the *span* of $\mathbf{x}$ in this simple case.

Orthogonal complement.

Consider the vector

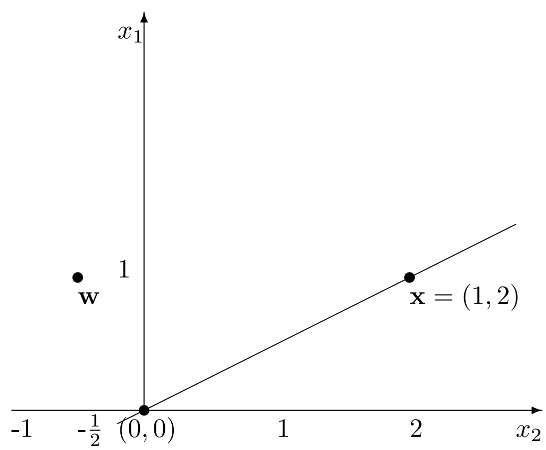
$$\mathbf{w} = (1, -\tfrac{1}{2})$$

Let us check that this is *orthogonal* to $\mathbf{x} = (1, 2)$.

$$\mathbf{w}'\mathbf{x} = (1, -\tfrac{1}{2})'(1, 2) = 1 \cdot 1 + (-\tfrac{1}{2}) \cdot 2 = 1 - 1 = 0$$

$\mathbf{w}$ is thus orthogonal to $\mathbf{x}$.

The vector $\mathbf{w}$ is a point in the same space as before

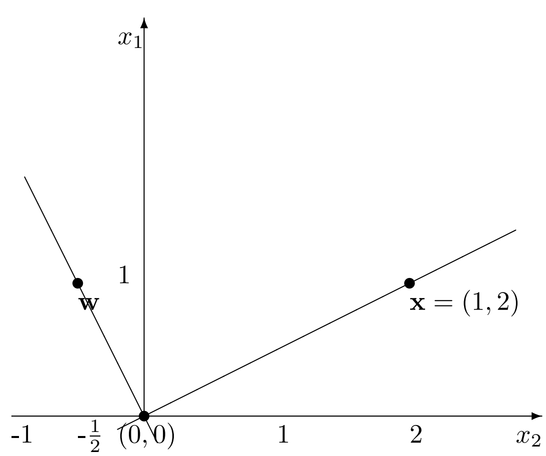


In fact, all vectors on a line $\gamma\mathbf{w}$ through the origin and $\mathbf{w}$ are orthogonal to $\mathbf{x}$. For example, $2\mathbf{w} = 2 \cdot (1, -\frac{1}{2}) = (2, -1)$ is also orthogonal to $\mathbf{x}$:

$$(2, -1)'(1, 2) = 2 \cdot 1 + (-1) \cdot 2 = 2 - 2 = 0$$

To see that all points on the line $\gamma\mathbf{w}$ is orthogonal:

$$(\gamma\mathbf{w}')(1, 2) = \gamma(1, -\tfrac{1}{2})'(1, 2) = \gamma(1 \cdot 1 - \tfrac{1}{2} \cdot 2) = \gamma \cdot 0 = 0$$



The line $\gamma\mathbf{w}$ is orthogonal to $\beta\mathbf{x}$, the two lines are at right angles.

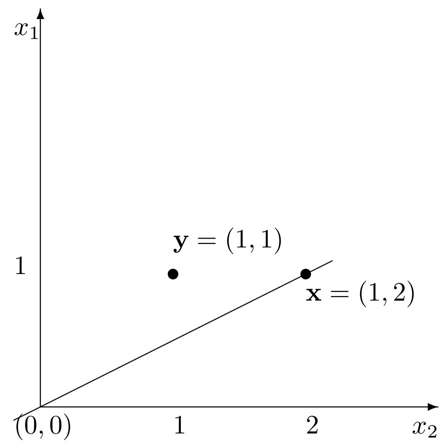
This picture is similar to DM fig 1.1, I only filled in some numbers.

Let us now consider regression in this picture.

We think of a vector $\mathbf{y}$, which we model as a linear function of $\mathbf{x}$,

$$\mathbf{y} = \mathbf{x}\beta$$

If there were no uncertainty in the picture, we could find β from one observation of $\mathbf{x}$ and $\mathbf{y}$, but the relation is not exact. Our problem is then to find a number $\hat{\beta}$ that "explains best" observations of $\mathbf{x}$ and $\mathbf{y}$.



The line in the figure is all possible ways of finding

$$\mathbf{x}\beta.$$

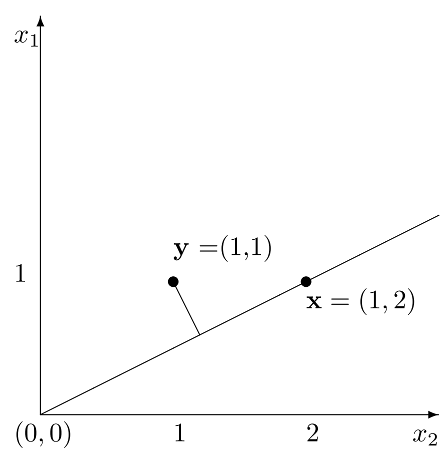
We want to choose the β, or the point on the line, that "best explains" the observed $\mathbf{y}$. If should make sense that one reasonable way of determining what "explains best" is to choose the $\mathbf{x}\beta$ that is *closest* to the observed $\mathbf{y}$. It is the criterion for what is "close" that determines our estimation method. In this picture we choose the usual (Euclidean) *distance* between $\mathbf{y}$ and $\mathbf{x}\beta$ as the criterion for fit.

We thus choose to minimize

$$\|\mathbf{y} - \mathbf{x}\beta\|,$$

with respect to β as our estimation procedure.

In the picture



We choose $\hat{\beta}$ as the point on $\mathbf{x}\beta$ with the minimum distance to $\mathbf{y}$. In the picture it is clear that this is also the point where the line from $\mathbf{y}$ is on a 90 degrees angle to $\mathbf{x}\beta$.

To find the minimum distance consider minimizing

$$(\mathbf{y} - \mathbf{x}\beta)' (\mathbf{y} - \mathbf{x}\beta)$$

This is an optimization problem, and we solve for the first order condition:

$$\frac{\partial}{\partial \beta} = -2\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta)$$

Set this equal to zero and solve for β.

$$\begin{aligned} \mathbf{x}'(\mathbf{y} - \mathbf{x}\beta) &= 0 \\ \rightarrow \quad \mathbf{x}'y - \mathbf{x}'\mathbf{x}\beta &= 0 \\ \rightarrow \quad \mathbf{x}'\mathbf{y} &= \mathbf{x}'\mathbf{x}\beta \\ \rightarrow \quad [\mathbf{x}'\mathbf{x}]^{-1} \mathbf{x}'\mathbf{y} &= \beta \end{aligned}$$

We have now found the famed OLS (ordinary least squares) estimate of a linear regression model:

$$\hat{\beta} = [\mathbf{x}'\mathbf{x}]^{-1} \mathbf{x}'\mathbf{y}$$

Let us find the actual numbers in our (silly) example,

$$\mathbf{x} = (1, 2)$$

$$\mathbf{y} = (1, 1)$$

What is the value of β?

$$\begin{aligned} \mathbf{x}'\mathbf{x} &= 5 \\ [\mathbf{x}'\mathbf{x}]^{-1} &= \frac{1}{5} \\ \mathbf{x}'\mathbf{y} &= 3 \\ \hat{\beta} &= \frac{3}{5} \end{aligned}$$

Thus, the "closest" point to $\mathbf{y}$ on the line $\mathbf{x}\beta$ is

$$\mathbf{x}\hat{\beta} = (1\ 2) \cdot \frac{3}{5} = \begin{pmatrix} \frac{3}{5} & \frac{6}{5} \end{pmatrix}$$

Many of the more complicated estimators of econometrics are variations of this general principle, they are solutions to a minimization problem.

Note that I still have said nothing about probability distributions and error terms, we are still only concerned with the consequences of minimizing a distance.

1.3 Normal equation

A note about the equation

$$\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta) = \mathbf{0}$$

This is called the *normal equation*.

Let us think about what this means in terms of the model

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{e}$$

Solve for $\mathbf{e}$:

$$\mathbf{e} = \mathbf{y} - \mathbf{x}\beta$$

In other words, the first order condition

$$\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta) = \mathbf{0}$$

can be written as

$$\mathbf{x}'\mathbf{e} = \mathbf{0}$$

or, the fitted errors are *orthogonal* to the data $\mathbf{x}$.

This can also be thought about intuitively

$$\mathbf{y} - \mathbf{x}\hat{\beta} = \hat{\mathbf{e}}$$

is the *unexplained* part of the regression. We have used the data $\mathbf{x}$ to find $\hat{\beta}$. What is left to explain is

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{x}\hat{\beta}$$

If we have used $\mathbf{x}$ optimally to find β, $\mathbf{x}$ should have nothing left to explain of $\hat{\mathbf{e}}$, that is, $\mathbf{x}$ is *orthogonal* to $\hat{\mathbf{e}}$, or

$$\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta) = \mathbf{x}'\mathbf{e} = \mathbf{0}$$

The normal equation

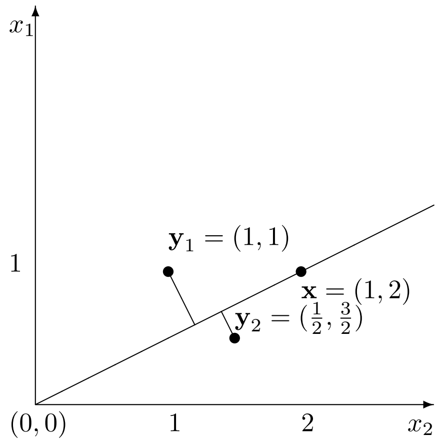
$$\mathbf{x}'(\mathbf{y} - \mathbf{x}\beta) = \mathbf{0}$$

has thus the interpretation that we use the information in $\mathbf{x}$ optimally to find $\hat{\beta}$

2 Measuring the fit of a regression.

The OLS estimate gives us a *method* to find an estimate of β. The next question is then concerned with the *quality* of this estimate. What we are concerned about is: How good is $\mathbf{x}\beta$ at predicting $\mathbf{y}$? We will answer this question much more formally later, but given these geometric pictures, you should think that there exists measures of the "fit" of a regression based on the *distance* between the observations $\mathbf{y}$ and the regression line $\mathbf{x}\beta$.

Consider the picture below, where we have two $\mathbf{y}$ observations, $\mathbf{y}_1$ and $\mathbf{y}_2$



Just looking at this picture and concluding that $\mathbf{y}_2$ is closer than $\mathbf{y}_1$ to $\mathbf{x}\beta$ may be misleading. We would like a measure of the distance that is not sensitive to *scaling*.

Such a measure is the "R-squared" of a regression. The R^2 can be explained in a number of ways. In geometric terms, look at the two points $\mathbf{y}_1$ and $\mathbf{y}_2$ above. If we want to compare the distance to $\mathbf{X}\beta$ of these two in a unit-free way, consider comparing the *angles* that the line from the origin to $\mathbf{y}_1$ and $\mathbf{y}_2$ forms with $\mathbf{X}\beta$.

Remember from linear vector spaces that the *angle* θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ was

$$\cos\theta = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|}$$

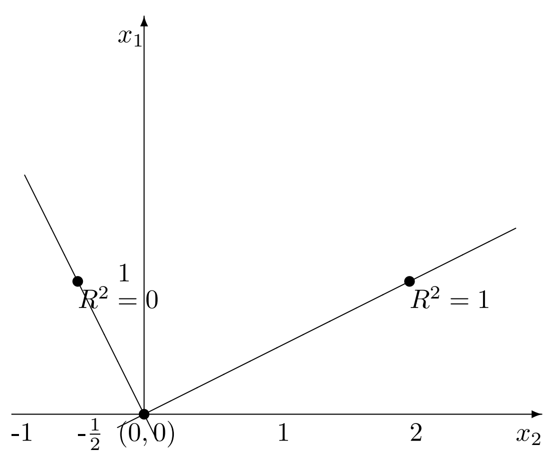
It turns out in terms of a OLS problem that the angle between $\mathbf{y}$ and the line $\mathbf{X}\beta$ is measured as

$$\cos\theta = \frac{\|\mathbf{y}'\hat{\beta}\|}{\|\mathbf{y}\|}$$

Taking the square of this, we find the (uncentered) R^2 of the regression,

$$R^2 = \cos^2\theta$$

Another way of interpreting R^2 is as the *fraction* of the errors explained by the regression. If $R^2 = 1$, then $\mathbf{y}$ is totally explained by the regression. If $R^2 = 0$, the regression explains nothing.



In this picture two cases are illustrated. $R^2 = 1$ means that $\mathbf{y}$ actually lies on the line $\mathbf{x}\beta$, everything is explained by $\hat{\beta}$. $R^2 = 0$ means that $\mathbf{y}$ is orthogonal to $\mathbf{X}\beta$, and lies in the orthogonal complement to $\mathbf{X}\beta$, nothing is explained by $\mathbf{X}\beta$.

Another way of interpreting R^2 is

$$R^2 = \frac{\text{Explained square sum of errors}}{\text{Total square sum of errors}}$$

R^2 is one when everything is explained, since the sum of squares is also our criterion function. If we were not estimating using least squares, R^2 has no meaning. R^2 is a useful measure of fit of a regression, but it should be used with care.

3 Geometry of multivariate regressions

Multivariate regressions is conceptually no different from univariate regressions, but the practicalities are (slightly) more complicated.

3.1 What multiple regressions are

The dependent variable y is now a function of several explanatory variables

$$y = a + b_1 x_1 + b_2 x_2 + ... + b_k x_k + e$$

The dependent variable y is now a linear function of k independent variables $x_1, x_2, ...x_k$.

The factors b_i have the same interpretation as b in the univariate regressions:

They measure the marginal effect of a change in one of the explanatory variables, *holding everyting else constant*

$$\frac{dy}{dx_i} = \frac{d(a + b_1 x_1 + b_2 x_2 + ...b_i x_i + ... + b_k x_k + e)}{dx_i} = b_i$$

This is written in matrix form

$$\tilde{y}_i = a + bx_{1i} + bx_{2i} + \cdots + b_k x_{ki} + \tilde{e}_i$$

$$\tilde{\mathbf{y}} = \mathbf{X}\mathbf{b}\tilde{\mathbf{e}}$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ ... & & & & \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} a \\ b_1 \\ \vdots \\ b_k \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

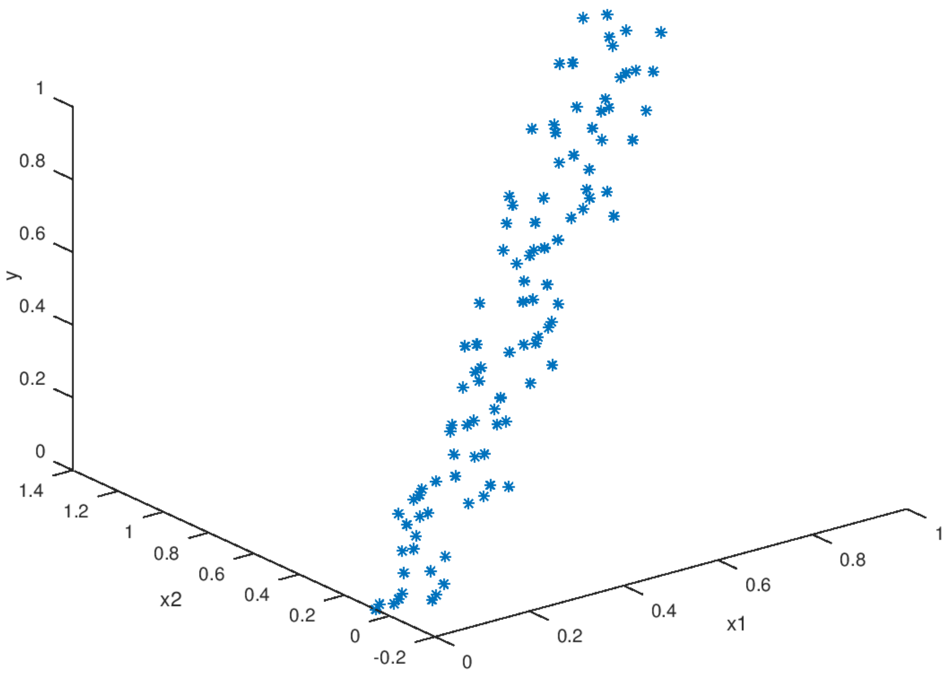
The regression is formulated as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

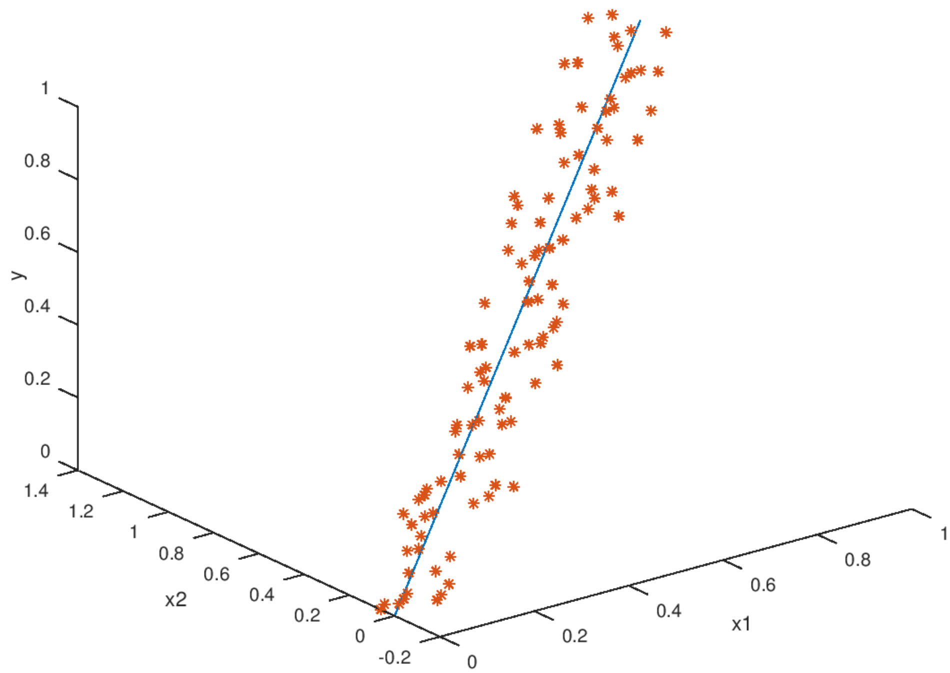
3.2 Geometrically

Regression line is a solution to minium distance problems - in several dimenstions.

Let us illustrate the idea. We consider a bivariate regression, where y is a function of two variables (x_1, x_2) We move to a three-dimensional picture, and data is then a bunch of points (x_1, x_2) in this space



Regression is then drawing the “best fitting” line in this space.



(For the interested, here is the code generating these pictures (in matlab)

```
x=[0:0.01:1];
y=x;
z=x;
y1=y+randn(1,101)*0.1;
plot3(x,y1,z,"*");
xlabel("x1");
ylabel("x2");
zlabel("y");
print("illustrating_regression_3d_points.eps","-depsc2");
plot3(x,y,z,x,y1,z,"*");
xlabel("x1");
ylabel("x2");
zlabel("y");
print("illustrating_regression_3d.eps","-depsc2");
```

3.3 Calculation of estimates

The minimum inference problem is

$$\mathbf{b} = \text{argmin}_{\mathbf{b}}(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

Calculation:

Step 1: the Normal Equation

$$\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{0}$$

Step 2: The analytical solution

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Sources Much of this material is taken from (Davidson and MacKinnon, 1993, 1.1, 1.2), which covers the basic material on how to do an OLS estimation.

(Davidson and MacKinnon, 1993, 1.6) discusses in more detail some of the mathematics of detection of outliers, and other problems in the data.

(Luenberger, 1969, Ch 4) is a good source on the optimization view of estimation.

4 Using a computer tool for Least Squares, octave

5 Using octave for OLS regressions.

We are now ready to do some estimation in octave. For now we are only interested in the mechanics of calculating estimates.

We start with a very simple univariate example.

Exercise 1.

At a large state university seven undergraduate students who are majoring in economics were randomly selected and surveyed. Two of the survey questions asked were:

- What was your grade-point average (GPA) in the preceding term?
- What was the average number of hours spent per week last term in the Orange and Brew?

The Orange and Brew is a favorite and only watering hole on campus. Using the data below, estimate with ordinary least squares the equation

$$G = \alpha + \beta H$$

where G is GPA and H is hours per week in Orange and Brew. What is the expected sign for β? Does the data support your expectations?

Student	GBP (G)	Hours per week in Orange and Brew (H)
1	3.6	3
2	2.2	15
3	3.1	8
4	3.5	9
5	2.7	12
6	2.6	12
7	3.9	4

Solution to Exercise 1.

```
Data = [1 , 3.6 , 3 ; \
2 , 2.2 , 15 ; \
3 , 3.1 , 8 ; \
4 , 3.5 , 9 ; \
5 , 2.7 , 12 ; \
6 , 2.6 , 12 ; \
7 , 3.9 , 4 ]
y=Data(:,2);
x=Data(:,3);
X=[ones(7,1),x]
b=ols(y,X)
b =
   4.25727
  -0.13017
```

Thus, we estimated the parameters as $\alpha = 4.25727$ and $\beta = -0.13017$.

The β is negative: The more of your time you spend drinking the worse your grades. (Did we need a regression to figure that out?)

Let next go to a multivariate setting, letting $\mathbf{y}$ be a function of a bivariate vector $\mathbf{X}$. We simulate a simple model.

Consider

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

We explain $\mathbf{y}$ as a linear function of $\mathbf{X}$. Suppose we have two explanatory variables. Then $\mathbf{b}$ is a 2×1 matrix. Let us simulate 100 observations of the model. In simulating, we need to add some noise $\mathbf{e}$ to the data, to avoid a perfect fit.

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

First fill in $\mathbf{X}$ by picking 100 random numbers between 0 and 1:

```
> X=rand(100,2);
```

Suppose

$$\mathbf{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

```
> b = [2;1]
b =
  2
  1
```

Now generate 100 observations of $\mathbf{y}$ by adding a small random error with mean zero to each observation.

```
> y = X*b + (randn(100,1)-0.5)*0.1;
```

Let us now estimate $\mathbf{b}$ in this model, using the OLS formula.

```
> bhat = inverse(X'*X)*X'*y
bhat =
  2.00744
  0.99017
```

And the result is close to the true $\mathbf{b} = [2, 1]'$.

`Octave` has actually a built in command that performs a OLS estimation. This is usually to prefer, it uses a more stable algorithm:

```
> ols(y,X)
ans =
  2.00744
  0.99017
```

The command is `ols(y,X)`, run an OLS estimation on the regression $\mathbf{y} = \mathbf{X}\mathbf{b}$. In this case it gives the same estimate of $\mathbf{b}$ as writing it in terms of the explicit formula, which it should. Any differences will be due to roundoff errors.

6 Some diagnostics on the OLS estimate.

Now, we have learned to run an OLS regression. Then the next question is what to do with these estimates. If all you want is to get the estimates, this is where you stop. You could for example use these predicted values for forecasting values of $\mathbf{y}$ for given $\mathbf{X}$ values.

```
> bhat
bhat =
  2.00744
  0.99017
> new_x = [1 1]
new_x =
  1  1
> forecast_y = new_x * bhat
forecast_y = 2.9976
```

But as a rule, you want to go further. We need to make further assumptions to provide specific probability statements about the accuracy of these estimates. But there is still a number of diagnostics we can perform without the need of that particular theory. Remember that what we are after is an investigation of *some* functional relationship

$$\mathbf{y} = f(\mathbf{X}) + \mathbf{e}$$

When we run an OLS regression we have made the assumption that $f(\cdot)$ is a *linear* function of $\mathbf{X}$.

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

There are some ways to investigate the correctness of this assumption by simple means. Given a OLS estimate $\widehat{\mathbf{b}}$, we can use these values of $\mathbf{b}$ to compare the fit of the *predicted* values of $\mathbf{y}$, $\hat{\mathbf{y}}$, to the observed values $\mathbf{y}$. A standard diagnostic of any model is to do a plot of the *residuals*, the difference between the *predicted* $\mathbf{y}$'s using the fitted parameter values, and the observed $\mathbf{y}$'s. It is often the case that plotting these residuals may reveal problems with the linearity assumption.

Let me illustrate this with some simulated data, similar to the simulations we did above.

Let us start by doing the same simulation as we did above, we had the true model to be

$$\mathbf{y} = \mathbf{X}\mathbf{b}$$

We fill the $\mathbf{X}$ matrix with random numbers, and generate the model by adding a normal error, before we run an OLS regression:

```
# file: leastsqr.m
# simulate simple OLS situation
#
X=rand(100,2);
b=[1;2];
e = 0.25 * randn(100,1);
y = X*b + e;
bhat = (inv(X'*X))*X'*y;
ehat = y - X*bhat;
```

After doing this, we have generated estimates $\hat{\mathbf{b}}$.

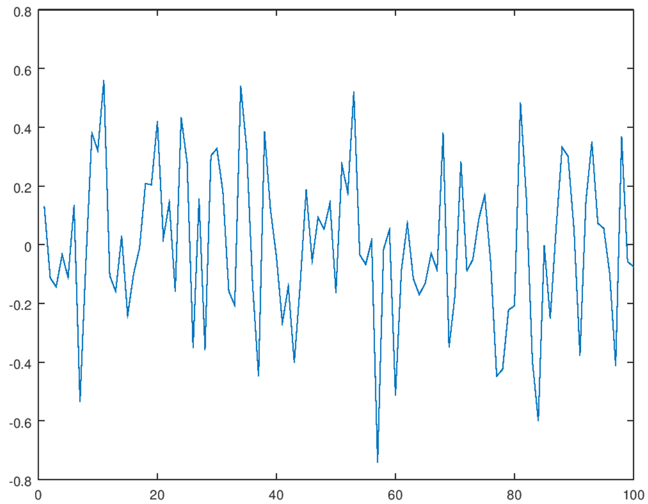
```
octave> bhat
bhat =
  0.98156
  1.99537
```

Since we have generated these data ourselves, we know that the residuals,

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\widehat{\mathbf{b}}$$

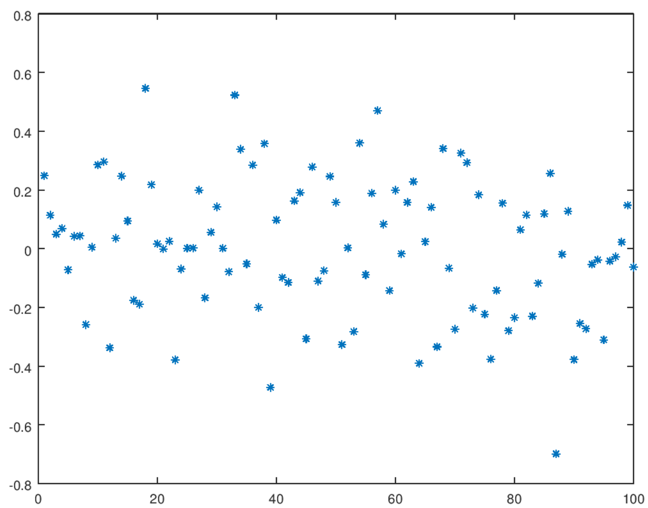
are normally distributed. To check that, we plot the residuals, and they should look random.

```
>> plot(ehat)
```



The lines in this plot are disturbing the view, let us plot the points only.

```
>> plot(ehat,"*");
```



The random character of the errors should be clearer in this picture.

7 Using residual plots to investigate misspecification

Consider a (univariate) model

$$y = a + bx + e$$

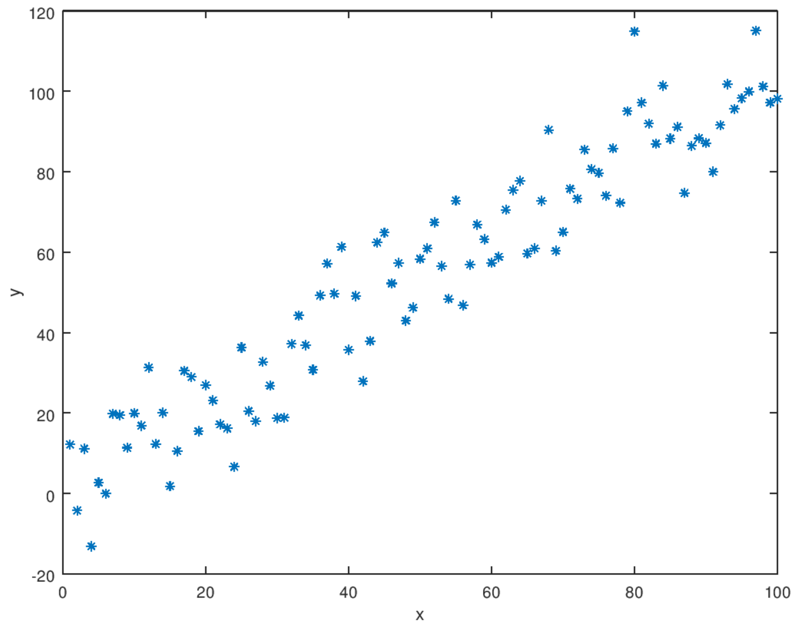
$$a = 1$$

$$b = 1$$

$$e \sim N(0, 10^2)$$

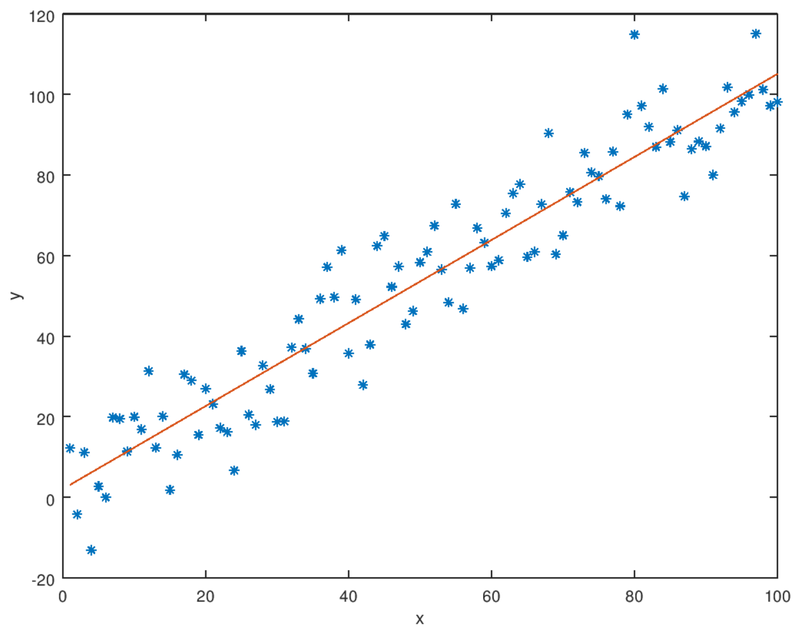
Simulate observations of this "true" model.

```
x=[1:100]';
X=[ones(100,1) x];
a=1;
b=1;
y=a*ones(100,1)+b*x+10*randn(100,1);
```

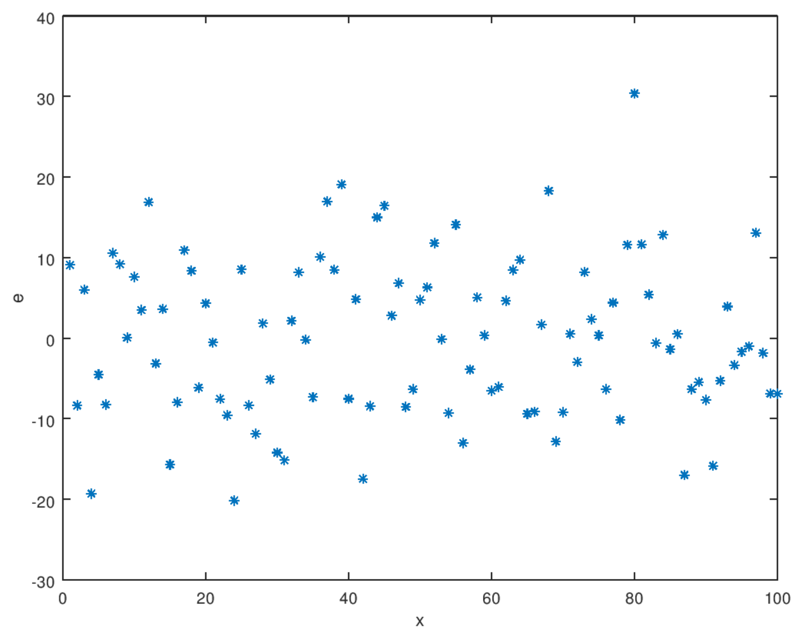


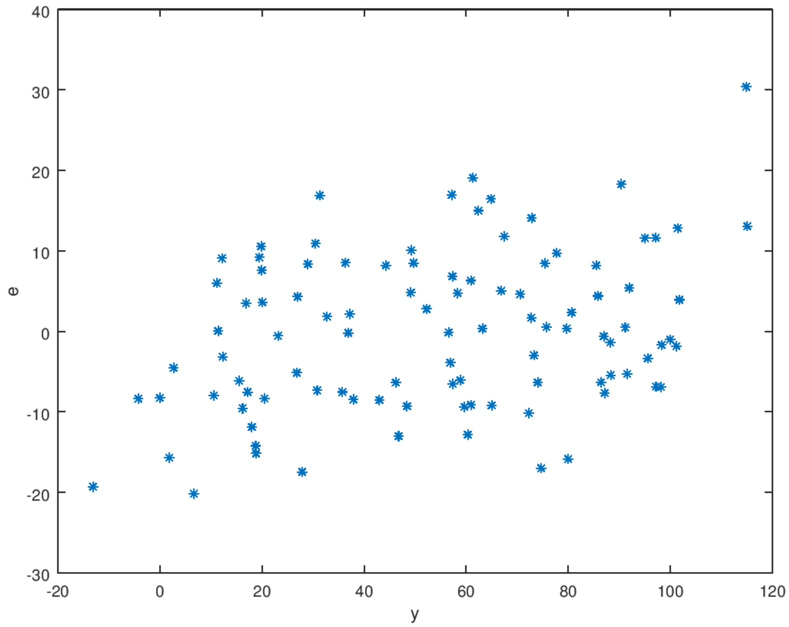
Fitted regression

```
b =
  -0.29247
   1.00990
```



What do the residuals look like?





40
30
20
10
0
-10
-20
-30
e
-20
0
20
40
60
80
100
120
y

Simulate the model

$$y = a + b\ln(x) + e$$

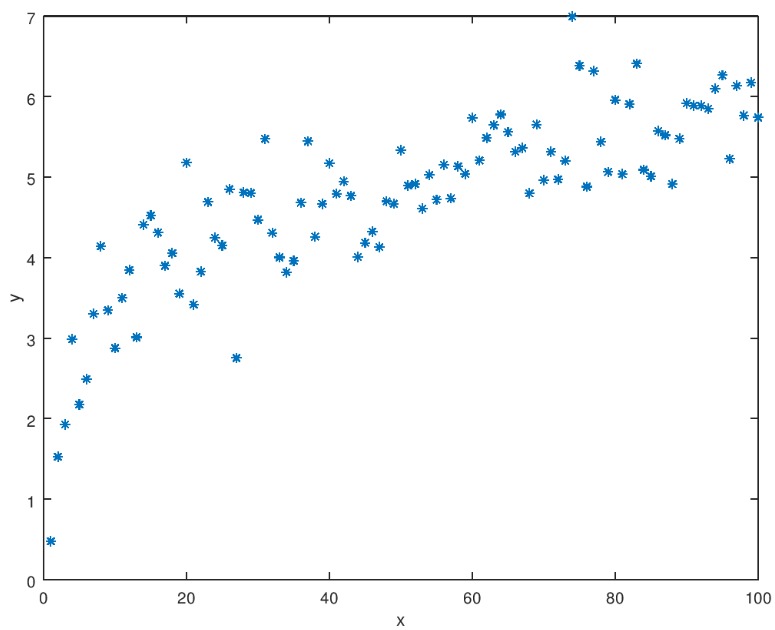
$$x = [1, 2, 3, ..., 100]$$

$$a = 1$$

$$b = 1$$

```
x=[1:100]';
a=1;
b=1;
y=a*ones(100,1)+b*log(x)+0.5*randn(100,1);
```

Plot observations

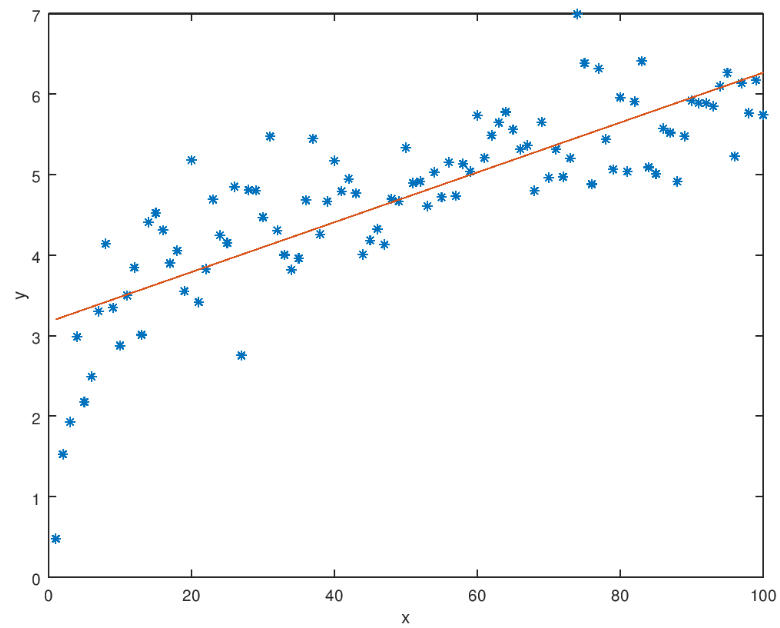


Fitted regression

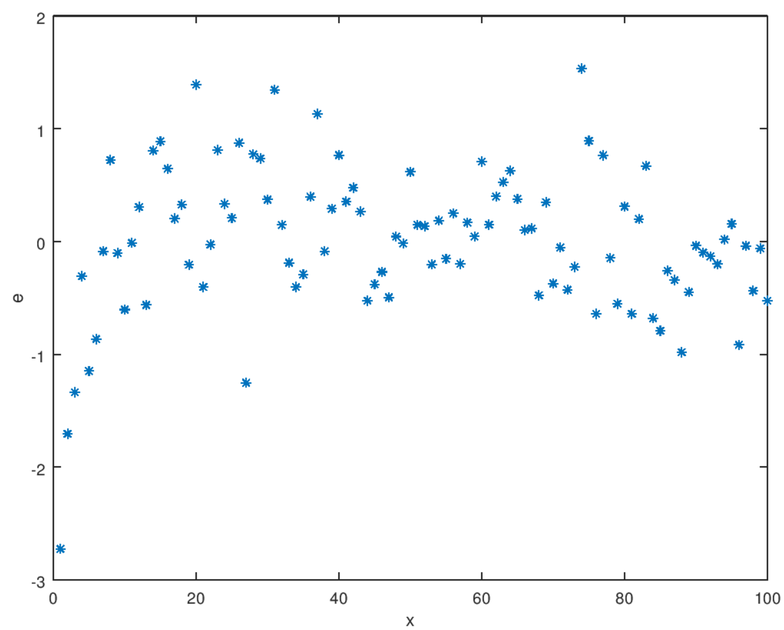
$$y = a + bx + e$$

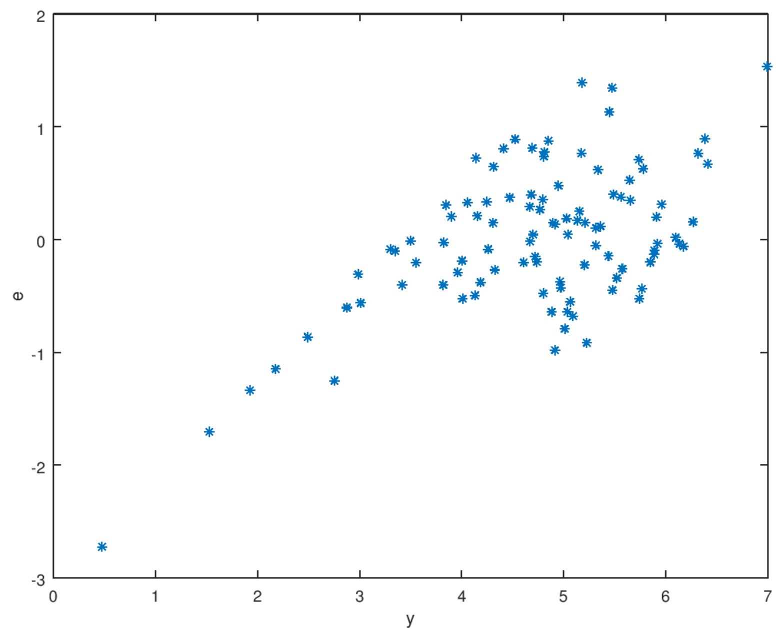
Get for example these results

```
b =
   3.102127
   0.044478
```



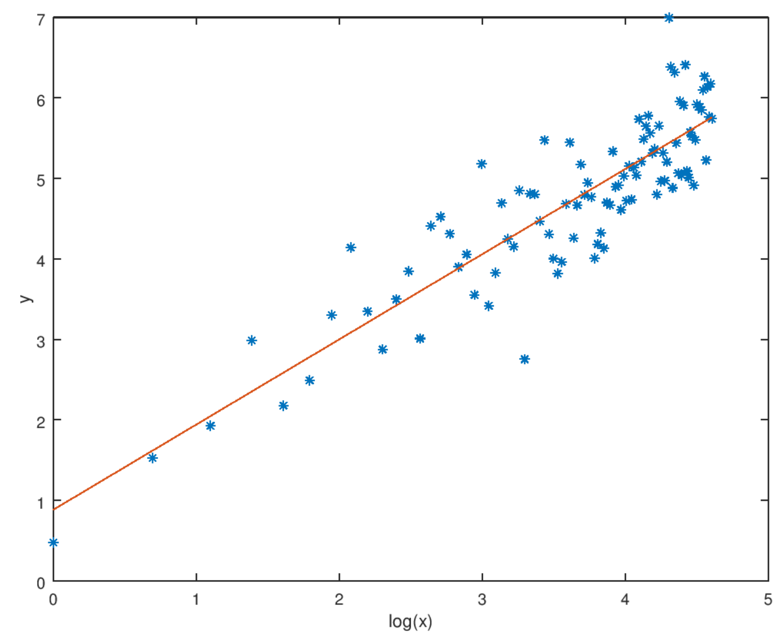
What do the residuals look like?



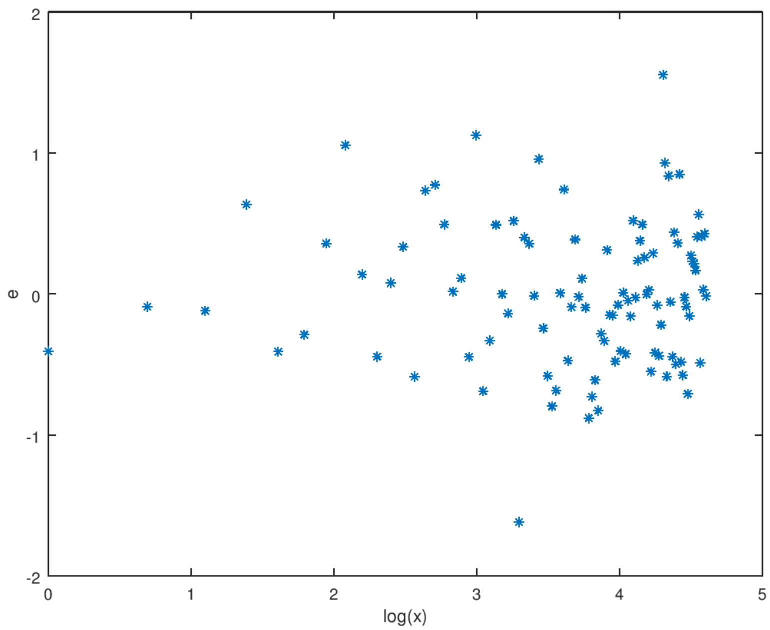


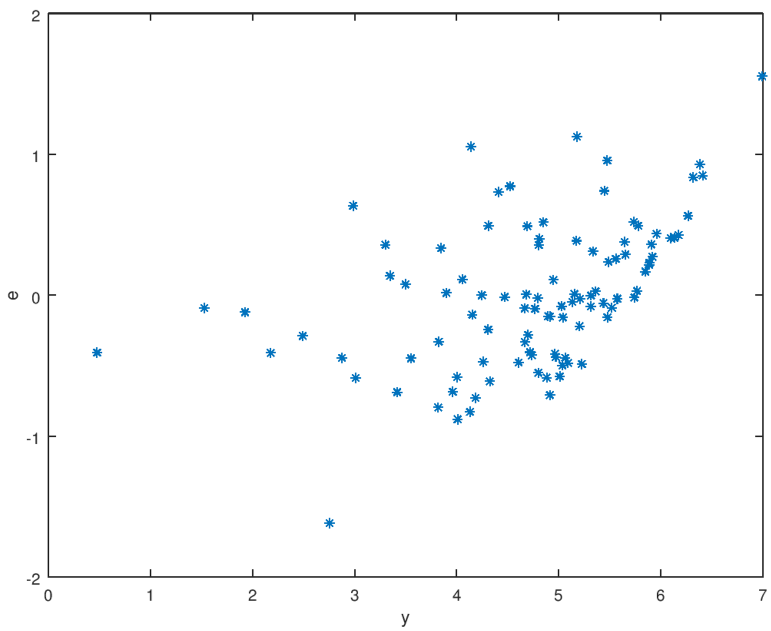
Aha, there is a problem.
Solution: define $x' = log(x)$ and run regression with this instead
Fitted regression
For example

```
bhat =
   0.23685
   1.22831
```

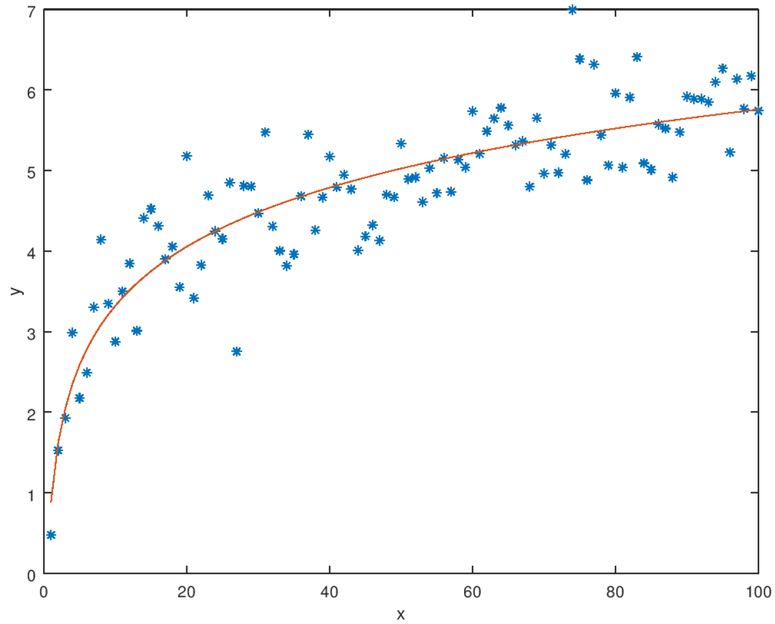


What do the residuals look like?





The nonlinear nature of the regression is clear when we plot it against x instead of $\ln(x)$.



7
6
5
4
3
2
1
0
y
0
20
40
60
80
100
x

Alternative nonlinear model

$$a = 1$$

$$b = 1$$

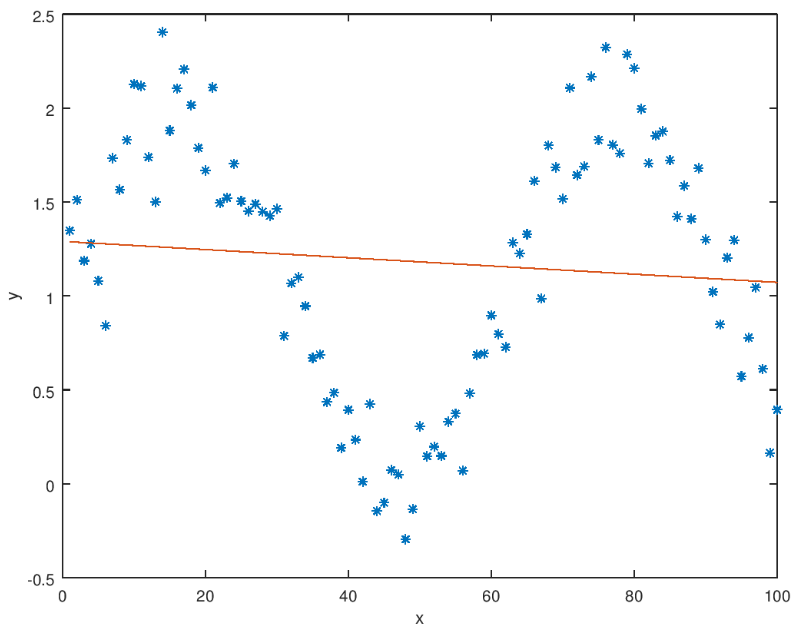
$$y = a + b\sin(0.1x) + e$$

```
a=1;
b=1;
y=a*ones(100,1)+b*sin(0.1*x)+0.25*randn(100,1);
```

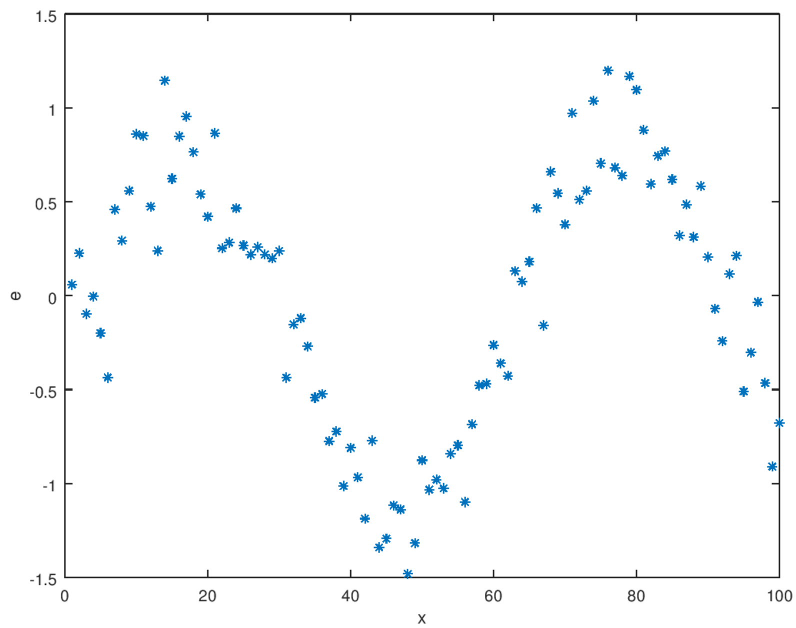
Fitted regression

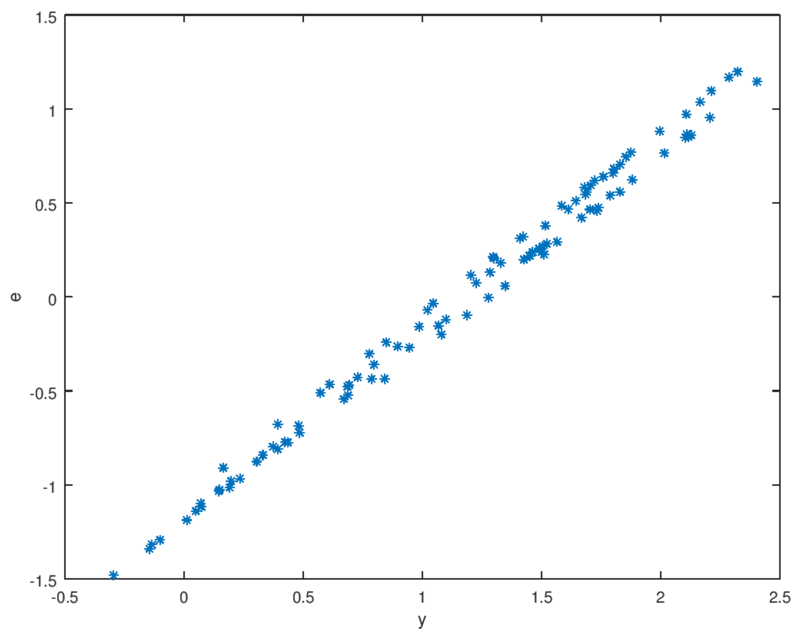
$$y = a + bx + e$$

```
b =
   1.4122e+00
  -5.9277e-04
```



What do the residuals look like?

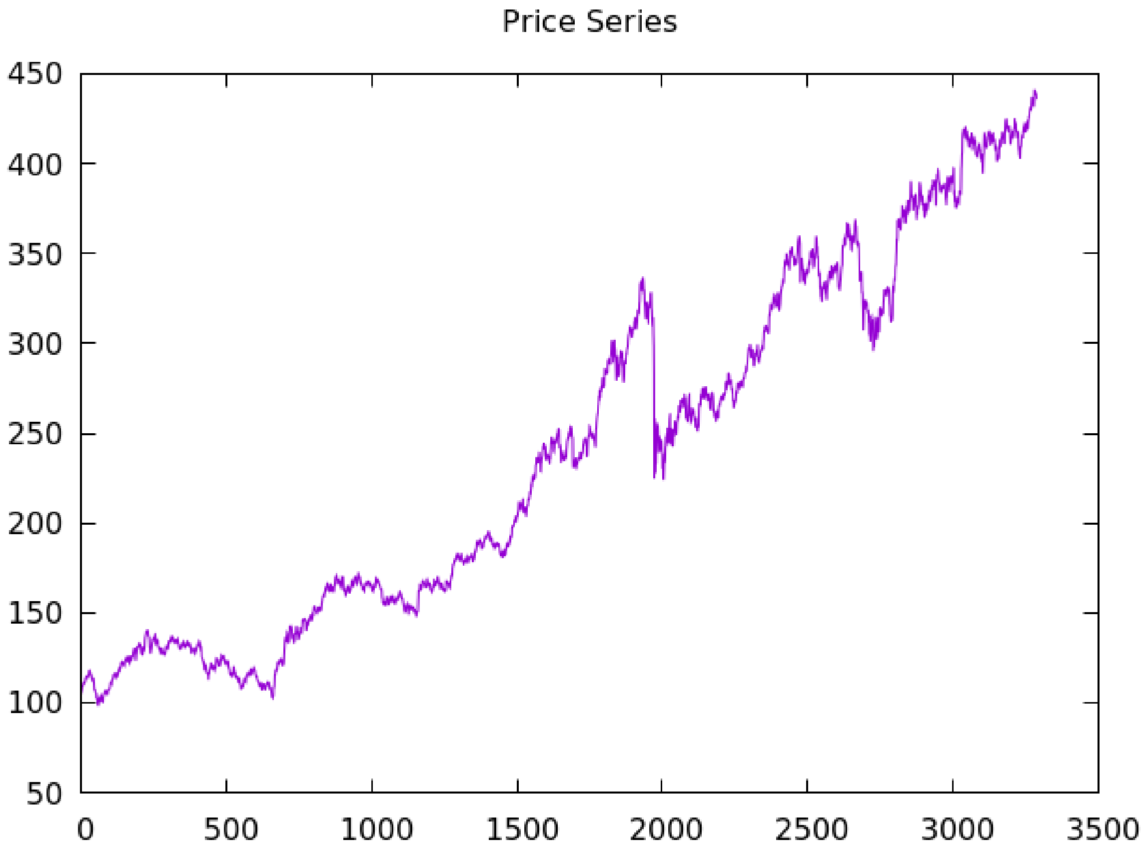




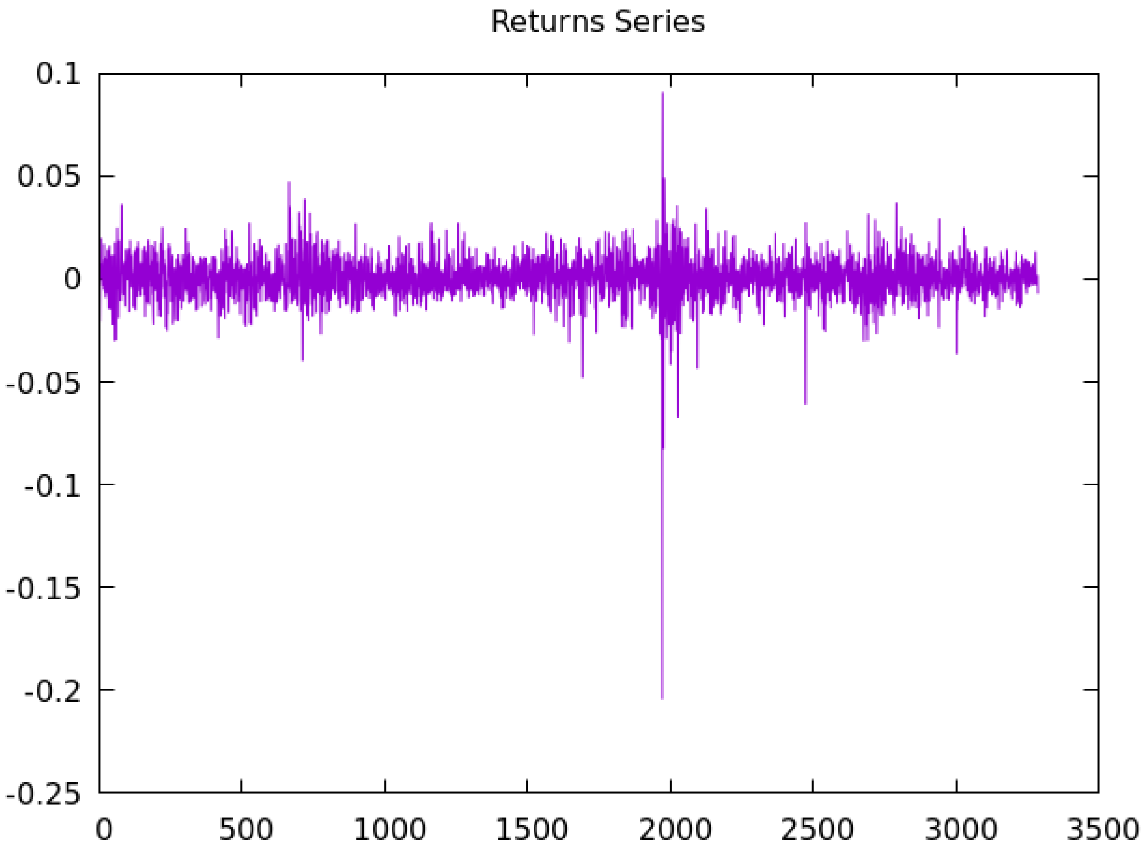
Aha, there is a problem.

8 Residuals outliers are not always errors...

One are not always justified in simply throwing out any observations considered an outlier. Consider the next picture, which is a time series of something...



There are what seems to be large jumps in the time series at a couple of points. If you were to difference this picture, as shown in the next figure,



the large jump observations look like outliers.

However, that observation is actually a true observation. This time series is the evolution of the Standard & Poors 500 stock price index for 1980 to 1992. The large drops in the index is the 20sep87 “crash”, and the “mini-crash” in 1989. What seems like “bad data” is not wrong data, it is just data that is hard to explain[1]

[1]We still can not explain the 1987 “crash.”

8.1 Use for regression: prediction

Once we have the estimated relationship, it can be used for for example prediction.

Exercise 2.

At a large state university seven undergraduate students who are majoring in economics were randomly selected and surveyed. Two of the survey questions asked were:

- What was your grade-point average (GPA) in the preceding term?
- What was the average number of hours spent per week last term in the Orange and Brew?

The Orange and Brew is a favorite and only watering hole on campus.

Using the data below, estimate with ordinary least squares the equation

$$G = \alpha + \beta H$$

where G is GPA and H is hours per week in Orange and Brew.

(The GPA is a numerical summary of grades with 4 as the largest number.)

What is the expected sign for β? Does the data support your expectations?

Student	GPA (G)	Hours per week in Orange and Brew (H)
1	3.6	3
2	2.2	15
3	3.1	8
4	3.5	9
5	2.7	12
6	2.6	12
7	3.9	4

Suppose that a freshman economics student has been spending 15 hours per week in the Orange and Brew during the first two weeks of class.

Predict his GPA for this year.

Solution to Exercise 2.

```
Data = [1 , 3.6 , 3 ; \
2 , 2.2 , 15 ; \
3 , 3.1 , 8 ; \
4 , 3.5 , 9 ; \
5 , 2.7 , 12 ; \
6 , 2.6 , 12 ; \
7 , 3.9 , 4 ]
y=Data(:,2);
x=Data(:,3);
X=[ones(7,1),x];
b=ols(y,X)
b =
   4.25727
  -0.13017
```

Thus, we estimated the parameters as $\alpha = 4.25727$ and $\beta = -0.13017$.

Now for prediction:

```
>> predicted=[1 15]*b
predicted =  2.3047
```

$$prediction = \alpha + \beta 15 = 4.25727 - 0.13017 \cdot 15 = 2.3047$$

What if number of hours is 4?

```
>> predicted=[1 4]*b
predicted =  3.7366
```

A bit better

9 Finance example: market model

Exercise 3.

You want to find the risk of the stock IBM. To do so you think the "market model"

$$r_{it} = \alpha_i + \beta_i r_{mt} + e_{it}$$

is a reasonable description of the risk return relationship. To estimate the parameters α_i and β_i you need a history of stock returns and index returns. Collect monthly returns for IBM and a broad based US stock market index, for example the S&P 500. Take data for 1995:1 to 2006:12.

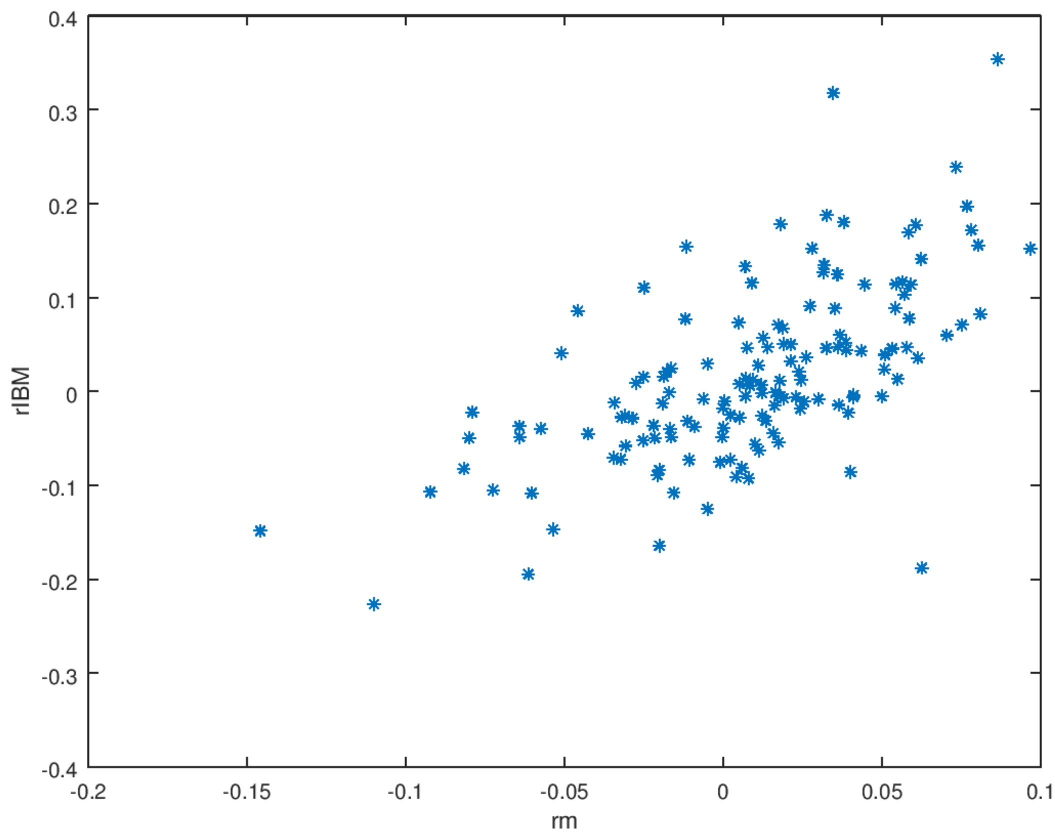
- Estimate the model.
- What is the R^2 in your estimation?

Use Matlab/Octave to implement the estimations.

Solution to Exercise 3.

The data is read into a couple of series

Plotting one against the other



Running the regression, get

```
>> r=ibm(:,2);
>> rm=sp500(:,2);
>> X=[ones(144,1) rm];
```

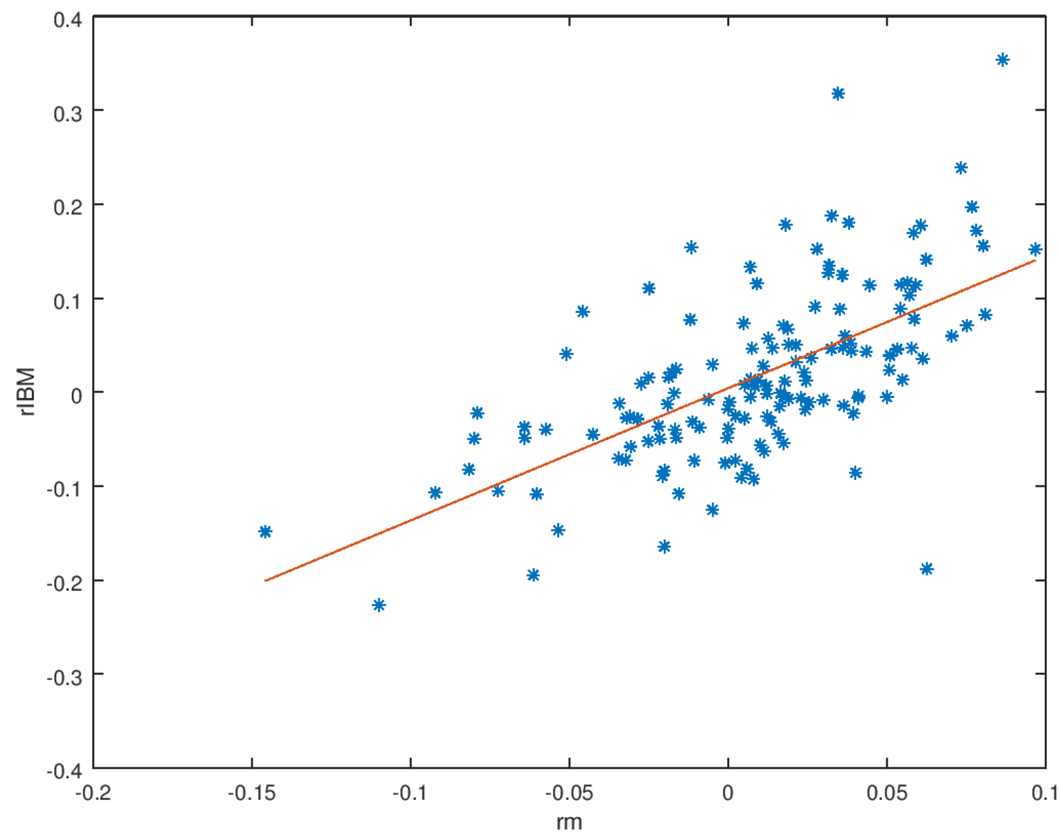
```
>> b=X\r
b =
   0.0041805
   1.4068184
```

We estimate the parameters as

$$\hat{a} = 0.0041805$$
$$\hat{b} = 1.4068184$$

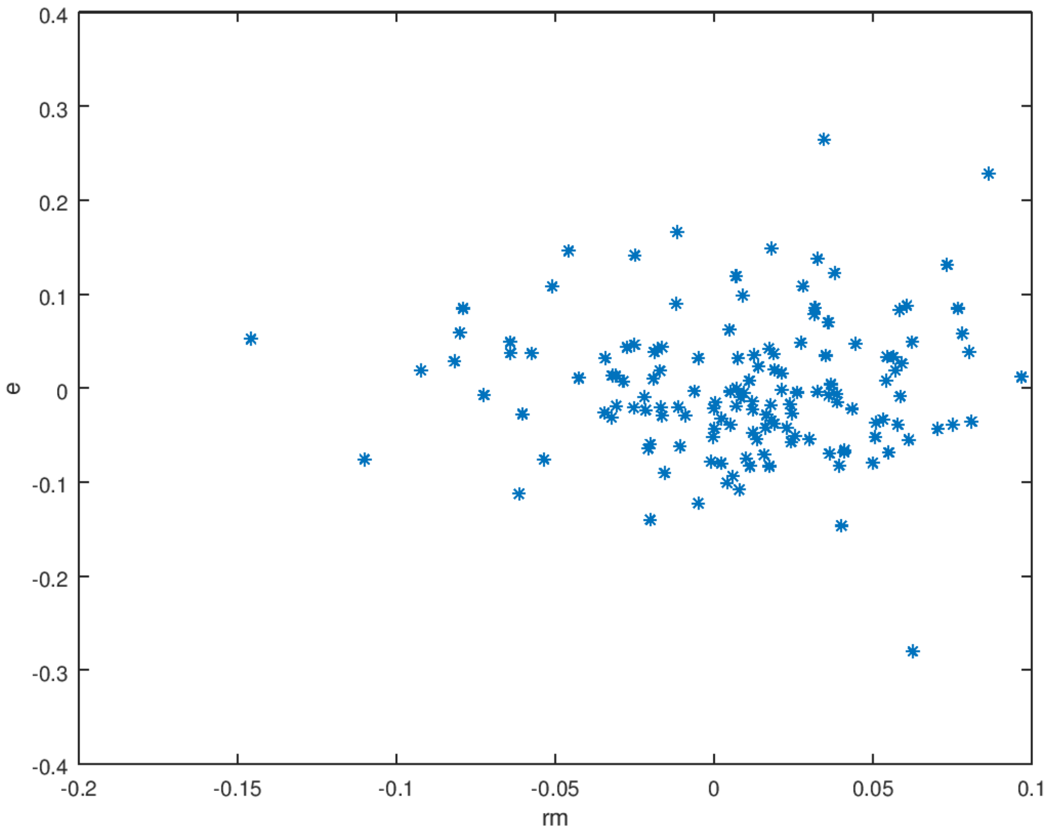
Next, plotting the observations and comparing it to the actual regrssion.

```
>> plot(rm,r,"*",rm,X*b)
```



Check for any obvious problems by calculating the residuals and plotting them against r_m:

```
>> plot(rm,e,"*");
```



Calculate the R^2

```
>> SSR=e'*e
SSR =  0.74944
>> TSS=(r-mean(r))'*(r-mean(r))
TSS =  1.2451
>> R2=1-SSR/TSS
R2 =  0.39811
```

The R^2 of the regression is 0.39811.

10 It is all optimization...

So far, all our analysis has actually solved an optimization problem to find the parameter estimates in a regression.

(minimum distance problem).

Remaining issue: Probability statements.

How can we evaluate an estimated coefficient – how "confident" are we that the true coefficient is close to what we have estimated.

To make such statements: Need additional assumptions.

References

Russel Davidson and James G MacKinnon. *Estimation and Interference in Econometrics*. Oxford University Press, 1993.

David G Luenberger. *Optimization by vector space methods*. Wiley, 1969.