Regression

Notation

- We need to extend our notation of the regression function to reflect the number of observations.
- As usual, we'll work with an iid random sample of *n* observations.
- If we use the subscript *i* to indicate a particular observation in our sample, our regression function with one independent variable is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{for } i = 1,2,\ldots,n$$

- So really we have *n* equations (one for each observation):

$$\begin{gathered} Y_1 = \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n \end{gathered}$$

Notice that the coefficients β_0 and β_1 are **the same** in each equation. The only thing that varies across equations is the data (Y_i, X_i) and the error ε_i.

Notation

- If we have more (say k) independent variables, then we need to extend our notation further.
- We could use a different letter for each variable (i.e., X, Z, W, etc.) but instead we usually just introduce another subscript on the X.
- So now we have two subscripts: one for the variable number (first subscript) and one for the observation number (second subscript).

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- What do the regression coefficients measure now? They are **partial derivatives, or marginal effects**. That is,

$$\beta_1 = \frac{\partial Y_i}{\partial X_{1i}} \quad \beta_2 = \frac{\partial Y_i}{\partial X_{2i}} \quad \cdots \quad \beta_k = \frac{\partial Y_i}{\partial X_{ki}}$$

So, β_1 measures the effect on Y_i of a one unit increase in X_{1i} **holding all the other variables X_{2i}, X_{3i}, ... , X_{ki} and ε_i constant.**

Data Generating Function

- Assume that the data *X* and *Y* satisfy (are generated by):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- The coefficients (β) and the errors (ε_i) are **not observed**.
- Sometimes our primary interest is the coefficients themselves
  - β_k measures the **marginal effect** of variable X_{ki} on the dependent variable Y_i.
- Sometimes we're more interested in predicting Y_i.
  - if we have sample estimates of the coefficients, we can calculate **predicted values:**

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki}$$

- In either case, we need a way to estimate the unknown β's.
  - That is, we need a way to compute $\hat{\beta}$'s from a sample of data
- It turns out there are lots of ways to estimate the β's (compute $\hat{\beta}$'s).
- By far the most common method is called **ordinary least squares (OLS)**.

Linearity

- There are **two** kinds of linearity present in the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- This regression function is **linear in *X*.**
  - *counter-example:* $Y = \beta_0 + \beta_1 X^2$
- This regression function is **linear in the coefficients** β_0 and β_1
  - *counter-example:* $Y = \beta_0 + X^{\beta}$
- Neither kind of linearity is necessary for estimation *in general*.
- We focus our attention mostly on what econometricians call the **linear regression model**.
- The linear regression model **requires** linearity in the coefficients, but **not** linearity in *X*.
  - When we say "linear regression model" we mean a model that is *linear in the coefficients*.

The Error Term

- Econometricians recognize that the regression function is never an **exact** representation of the relationship between dependent and independent variables.
  - e.g., there is no exact relationship between income (Y) and education, gender, etc., because of things like luck
- There is **always** some variation in Y that cannot be explained by the model.
- There are many possible reasons: there might be "important" explanatory variables that we leave out of the model; we might have the wrong functional form (f), variables might be measured with error, or maybe there's just some randomness in outcomes.
- These are all sources of **error**. To reflect these kinds of error, we include a **stochastic (random) error term** in the model.
- The error term reflects all the variation in Y that cannot be explained by X.
- Usually, we use epsilon (ε) to represent the error term.

More About the Error Term

- It is helpful to think of the model as having two components:

  1. a *deterministic* (non-random) component

  2. a *stochastic* (random) component ε

- Basically, we are decomposing Y into the part that we can explain using X and the part that we cannot explain using X (i.e., the error ε)

- We usually assume things about the error ε.
- I want to assume $E[\varepsilon_i|X_i]=0$ and $E[\varepsilon_i|\varepsilon_j]=0$ for a while.
  - This is overly strong, but is okay for a while.

Conditional Expectation Functions

- The mean of the conditional distribution of Y given X is called the **conditional expectation** (or **conditional mean**) **of** $\boldsymbol{Y}$ **given** $\boldsymbol{X}$.
- It's the expected value of Y, given that X takes a particular value.
- From Review 1, in general, it is computed just like a regular (unconditional) expectation, but uses the conditional distribution instead of the marginal.
  - If Y takes one of k possible values $y_1, y_2, \ldots, y_k$ then:

$$E(Y \mid X = x) = \sum_{i=1}^{k} y_i \Pr(Y = y_i \mid X = x)$$

- The conditional expectation of Y given the linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

- is

$$E[Y_i \mid X_1 = X_{1i}, \ldots, X_k = X_{ki}] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki}$$

- because $E[\varepsilon_i \mid X_1 = X_{1i}, \ldots, X_k = X_{ki}] = 0$

Marginal Effect

- Given these assumptions on the error term, we can strengthen the meaning of the coefficient: it is the marginal effect of X on Y, regardless of the value of ε.
- This is because the derivative of Y with respect to X, which *could* depend on the value of ε through the dependence of ε on X, does not depend on the value of ε.
  - Prove this via application of the chain rule.

What is Known, What is Unknown, and What is Assumed

- It is useful to summarize what is known, what is unknown, and what is hypothesized.
- **Known:** Y_i and X_{1i} , X_{2i} , ... , X_{ki} (the data)
- **Unknown:** β_0 , β_1 , β_2 , ... , β_k and ε_i (the coefficients and errors)
- **Hypothesized:** the form of the regression function, e.g.,

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki}$$

- We use the observed data to learn about the unknowns (coefficients and errors), and then we can test the hypothesized form of the regression function.
- We can hope to learn a lot about the βs because they are the same for each observation.
- We can't hope to learn much about the ε_i because there is only one observation on each of them.

Even Simpler Regression

- Suppose we have a linear regression model with one independent variable and NO INTERCEPT: $Y_i = \beta X_i + \varepsilon_i$
- Suppose also that

$$E[\varepsilon_i] = 0 \text{ and } E[(\varepsilon_i)^2] = \sigma^2 \text{ and } E[(\varepsilon_i \varepsilon_j)] = 0 \text{ for all } i, j$$

- Now, define an estimator as the number $\hat{\beta}$ that minimises the sum of the squared prediction error

$$e_i = Y_i - \hat{\beta} X_i$$

- $\underset{\hat{\beta}}{\text{Min}}$ $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta} X_i\right)^2 = \sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n} \left(2 Y_i \hat{\beta} X_i\right) + \sum_{i=1}^{n} \left(\hat{\beta} X_i\right)^2$

Minimisation

- The squared Y leading term doesn't have $\hat{\beta}$
- Min $\underset{\hat{\beta}}{}$ $$-\sum_{i=1}^{n}\left(2Y_i\hat{\beta}X_i\right)+\sum_{i=1}^{n}\left(\hat{\beta}X_i\right)^2$$
- First-Order Condition

$$-2\sum_{i=1}^{n}\left(X_iY_i\right)+2\hat{\beta}\sum_{i=1}^{n}\left(X_i^{\,2}\right)=0$$

$$-\sum_{i=1}^{n}\left(X_iY_i\right)+\hat{\beta}\sum_{i=1}^{n}\left(X_i^{\,2}\right)=0$$

$$\hat{\beta}\sum_{i=1}^{n}\left(X_i^{\,2}\right)=\sum_{i=1}^{n}\left(Y_iX_i\right)$$

$$\hat{\beta}=\frac{\sum_{i=1}^{n}X_iY_i}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}$$

OLS Coefficients are Sample Means

- The estimated coefficient is a weighted average of the *Y*'s:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} \left(X_i^2 \right)} = \sum_{i=1}^{n} w_i Y_i$$

$$w_i = \frac{X_i}{\sum_{i=1}^{n} \left(X_i^2 \right)}$$

- It is a function of the data (a special kind of sample mean), and so it is a *statistic*.
- It can be used to estimate something we are interested in: the population value of β
- Since it is a statistic, it has a sampling distribution that we can evaluate for bias and variance.

Bias

- Pretend *X* is not random. Remember assumptions from above:

$$\boxed{Y_i = \beta X_i + \varepsilon_i}$$

- $$\boxed{E[\varepsilon_i] = 0 \ and \ E[(\varepsilon_i)^2] = \sigma^2 \ and \ E[(\varepsilon_i \varepsilon_j)] = 0 \ for\, all\, i, j}$$

- Substitute into the estimator and take an expectation:

$$E\left[\hat{\beta}\right] = E\left[\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\right] = E\left[\frac{\sum_{i=1}^{n} X_i\left(\beta X_i + \varepsilon_i\right)}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\right]$$

$$= \beta E\left[\frac{\sum_{i=1}^{n} X_i\left(X_i\right)}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\right] + E\left[\frac{\sum_{i=1}^{n} X_i \varepsilon_i}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\right] = \beta + 0 = \beta$$

Variance

- $$V\left[\hat{\beta}\right]=E\left[\left(\hat{\beta}-E\left[\hat{\beta}\right]\right)^2\right]=E\left[\left(\frac{\sum_{i=1}^{n}X_i\varepsilon_i}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\right)^2\right]=\frac{1}{\left(\sum_{i=1}^{n}\left(X_i^{\,2}\right)\right)^2}E\left[\left(\sum_{i=1}^{n}X_i\varepsilon_i\right)^2\right]$$

$$=\frac{1}{\left(\sum_{i=1}^{n}\left(X_i^{\,2}\right)\right)^2}E\left[X_1X_1\varepsilon_1\varepsilon_1+X_1X_2\varepsilon_1\varepsilon_2+\ldots+X_{n-1}X_n\varepsilon_{n-1}\varepsilon_n+X_nX_n\varepsilon_n\varepsilon_n\right]$$

$$=\frac{1}{\left(\sum_{i=1}^{n}\left(X_i^{\,2}\right)\right)^2}E\left[\sum_{i=1}^{n}\left(X_i\right)^2\left(\varepsilon_i\right)^2\right]=\frac{\sum_{i=1}^{n}\left(X_i\right)^2}{\left(\sum_{i=1}^{n}\left(X_i^{\,2}\right)\right)^2}E\left[\left(\varepsilon_i\right)^2\right]=\frac{1}{\sum_{i=1}^{n}\left(X_i^{\,2}\right)}\sigma^2$$

Simple Linear Regression

- Suppose now that we have a linear regression model with one independent variable and an intercept: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
- Suppose also that

$$E[\varepsilon_i] = 0 \ and \ E[(\varepsilon_i)^2] = \sigma^2 \ and \ E[(\varepsilon_i \varepsilon_j)] = 0 \ for \ all \ i, j$$

- Now, define an estimator as the number $\hat{\beta}$ that minimises the sum of the squared prediction error

$$e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

- $\underset{\hat{\beta}}{\text{Min}}$ $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2$

Minimisation

First-Order Condition, apply the chain rule to the square function:

$$2\sum_{i=1}^{n}\frac{\partial\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)}{\partial\hat{\beta}_0}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)=0$$

$$2\sum_{i=1}^{n}\frac{\partial\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)}{\partial\hat{\beta}_1}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)=0$$

Differentiate again:

$$-2\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)=2\sum_{i=1}^{n}e_i=0$$

$$-2\sum_{i=1}^{n}X_i\left(Y_i-\hat{\beta}_0-\hat{\beta}_1 X_i\right)=2\sum_{i=1}^{n}X_i e_i=0$$

Mean of residuals is zero; Covariance of residuals and *X* is zero.

Minimisation

$$\hat{\beta}_0 :$$

$$-2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\overline{Y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{X} = 0$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1 :$$

$$-2\sum_{i=1}^{n} X_i \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\sum_{i=1}^{n} X_i \left(Y_i - \left(\overline{Y} - \hat{\beta}_1 \overline{X}\right) - \hat{\beta}_1 X_i\right) = 0$$

$$\sum_{i=1}^{n} X_i \left(\left(Y_i - \overline{Y}\right) - \hat{\beta}_1 \left(X_i - \overline{X}\right)\right) = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i \left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n} X_i \left(X_i - \overline{X}\right)}$$

$$FOC \ for \ \hat{\beta}_0 \ implies :$$

$$-2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\overline{X}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\sum_{i=1}^{n}\overline{X}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$so \ we \ can \ write \ \hat{\beta}_1 :$$

$$-2\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

OLS Coefficients are Sample Means

- The estimated coefficients are weighted averages of the *Y*'s:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = \sum_{i=1}^{n}\left(\frac{\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} - \frac{1}{n}\right)Y_i$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X} = \sum_{i=1}^{n}\left(\frac{1}{n} - \overline{X}\left(\frac{\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} - \frac{1}{n}\right)\right)Y_i$$

- It is a function of the data (a special kind of sample mean), and so it is a *statistic*.
- It can be used to estimate something we are interested in: the population value of β
- Since it is a statistic, it has a sampling distribution that we can evaluate for bias and variance.

OLS estimator is unbiased

$$E\left[\hat{\beta}_1\right] = E\left[\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right] = E\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(\beta_0 + \beta_1 X_i + \varepsilon_i - \overline{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right]$$

$$= E\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(\beta_0 + \beta_1 X_i + \varepsilon_i - \beta_0 - \beta_1 \overline{X} - \overline{\varepsilon}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right]$$

$$= \beta_1 E\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\right] + E\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\varepsilon_i}{\sum_{i=1}^{n}\left(X_i{}^2\right)}\right] - E\left[\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\overline{\varepsilon}}{\sum_{i=1}^{n}\left(X_i{}^2\right)}\right]$$

$$= \beta_1 + 0 + 0 = \beta_1$$

Variance of OLS estimator

- Variance is more cumbersome to work out by hand, so I won't do it:
- Top looks like the
- "even simpler" model.

$$Var(\beta_1) = \frac{1}{\left(\sum_{i=1}^{n} X_i^{\;2}\right) - n\overline{X}^2}\sigma^2$$

$$= \frac{1}{n\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i^{\;2}\right) - \overline{X}^2\right)}\sigma^2$$

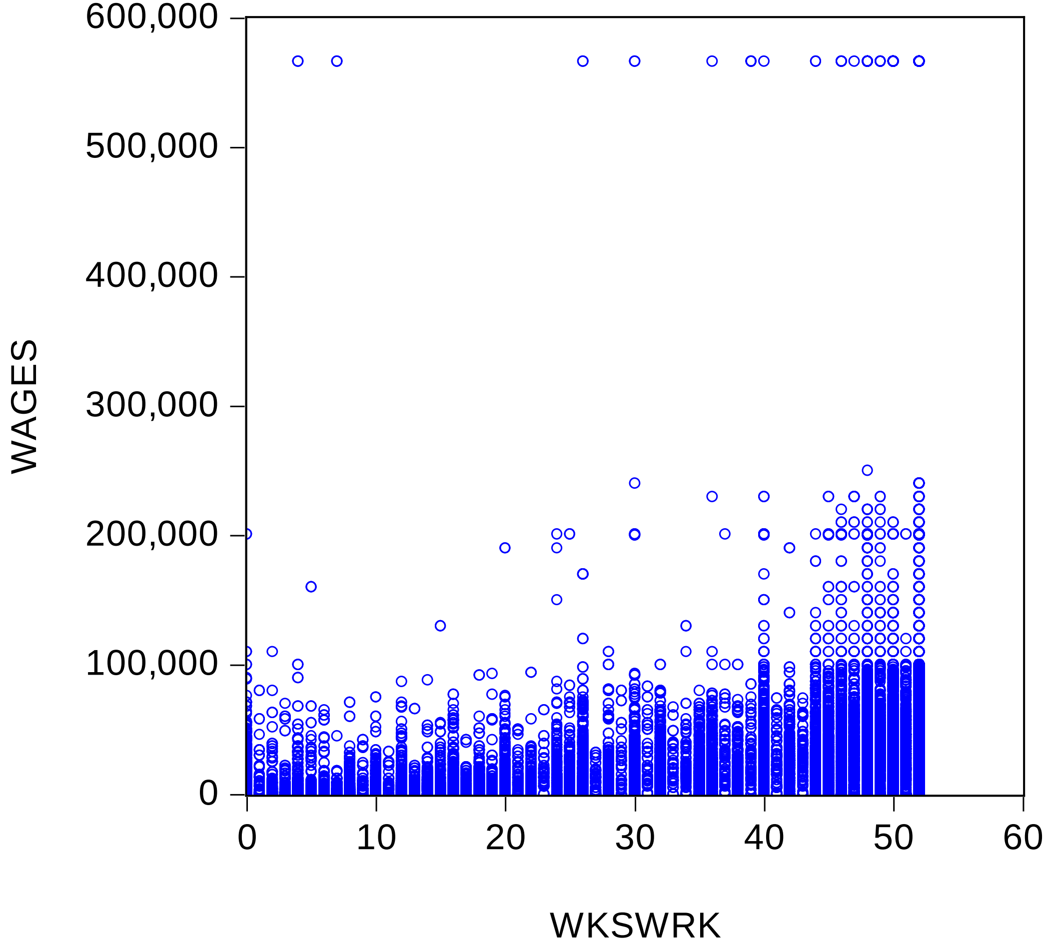
$$= \frac{1}{n\widehat{Var}(X)}\sigma^2$$

- Where V-hat is the
- sample variance of *X*
- $V(X)=E[X^2]-(E[X])^2$

In Practise…

- Knowing the summation formulas for OLS estimates is useful for understanding how OLS estimation works.
  - once we add more than one independent variable, these summation formulas become cumbersome
  - In practice, we never do least squares calculations by hand (that's what computers are for)
- In fact, doing least squares regression in EViews is a piece of cake – time for an example.

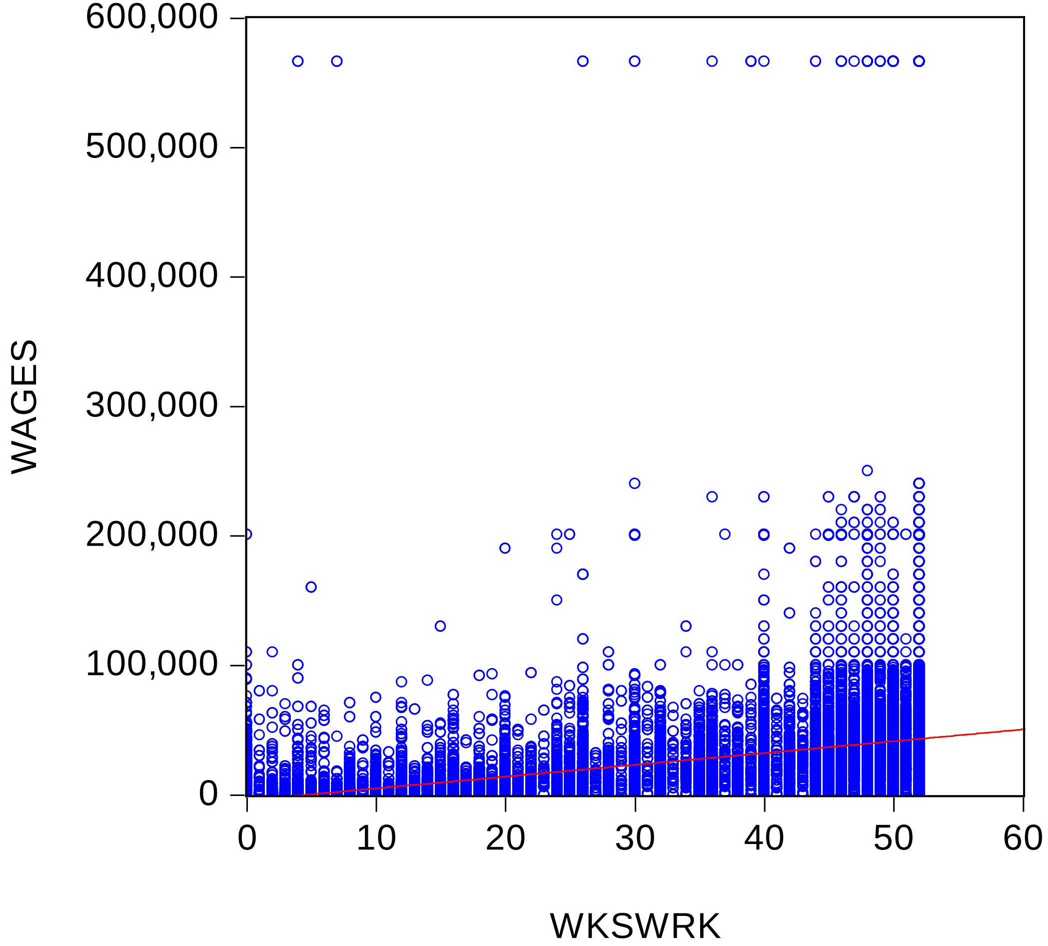
Example: Earnings and Weeks Worked



Example: Earnings and Weeks Worked

- Suppose we are interested in how weeks of work relate to earnings.
  - our dependent variable (Y_i) will be WAGES
  - our independent variable (X_i) will be WKSWRK
- After opening the EViews workfile, there are two ways to set up the equation:
  1. select WAGES and then WKSWRK (the order is important), then right-click one of the selected objects, and OPEN -> AS EQUATION , with an `IF WAGES<8000000 and WKSWRK<99` in the sample box (to get rid of both types of missing)
  or
  2. QUICK -> ESTIMATE EQUATION and then in the EQUATION SPECIFICATION dialog box, type:
  `wages c wkswrk`
  (the first variable in the list is the dependent variable, the remaining variables are the independent variables including the intercept `c`) and
- `if wages<8000000 and wkswrk<99 and agegrp>5 and agegrp<17` in the sample box
- You'll see a drop down box for the estimation METHOD, and notice that least squares (LS) is the default. Click OK.
- It's as easy as that. Print the output to rich text (rtf). Your results should look like the next slide ...

Data and Regression Line



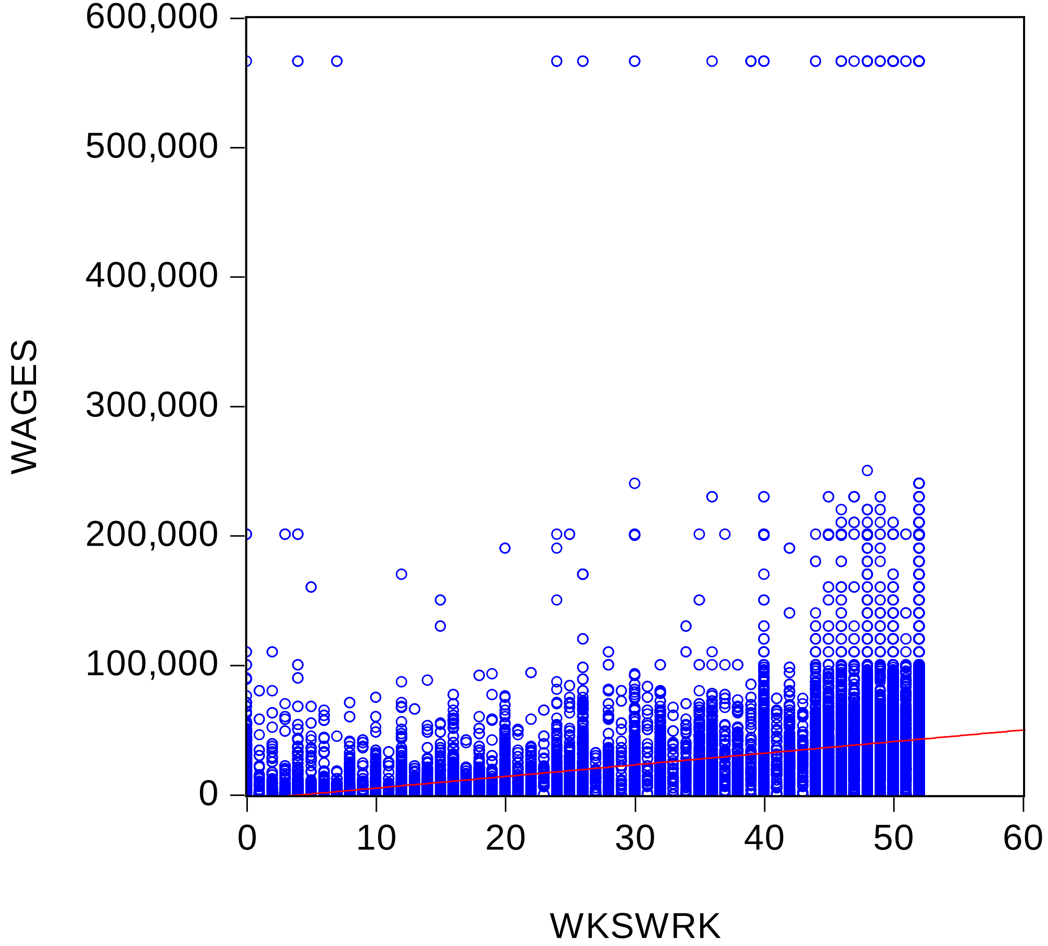
Eviews Regression Output

- Dependent Variable: WAGES
- Method: Least Squares
- Date: 10/19/10 Time: 10:18
- Sample: 1 56529 IF WAGES<8000000 AND WKSWRK<99 AND
- AGEGRP>5 AND AGEGRP<17
- Included observations: 32765

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3910.838	746.7136	-5.237401	0.0000
WKSWRK	910.5101	16.76971	54.29493	0.0000

R-squared	0.082550	Mean dependent var	34029.35
Adjusted R-squared	0.082522	S.D. dependent var	49743.57
S.E. of regression	47646.91	Akaike info criterion	24.38108
Sum squared resid	7.44E+13	Schwarz criterion	24.38160
Log likelihood	-399421.1	Hannan-Quinn criter.	24.38125
F-statistic	2947.939	Durbin-Watson stat	2.110695
Prob(F-statistic)	0.000000		

The Data and Regression Line



Low Variance is Good

- Low variance gives you a nice accurate picture of where the coefficient probably is.
- Low variance gives you a test statistic with high power (if the null is false, you'll probably reject).

How Do You Get Low Variance?

- The OLS estimator is unbiased, so it centers on the right thing.
- Its variance $Var(\beta_1) = \frac{1}{n\widehat{Var(X)}}\sigma^2$ has 3 pieces:
- N
- V(X)
- sigma-squared
- (draw them all)

Eviews Regression Output Again

- Dependent Variable: WAGES
- Method: Least Squares
- Date: 10/19/10 Time: 10:18
- Sample: 1 56529 IF WAGES<8000000 AND WKSWRK<99 AND
- AGEGRP>5 AND AGEGRP<17
- Included observations: 32765

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3910.838	746.7136	-5.237401	0.0000
WKSWRK	910.5101	16.76971	54.29493	0.0000

R-squared	0.082550	Mean dependent var	34029.35
Adjusted R-squared	0.082522	S.D. dependent var	49743.57
S.E. of regression	47646.91	Akaike info criterion	24.38108
Sum squared resid	7.44E+13	Schwarz criterion	24.38160
Log likelihood	-399421.1	Hannan-Quinn criter.	24.38125
F-statistic	2947.939	Durbin-Watson stat	2.110695
Prob(F-statistic)	0.000000		

Is That Coefficient Interesting?

- Weeks worked is correlated with a lot of things, eg, the young and old don’t work as many weeks.
- Add age
  - Recode agegrp into age:
    - series age=@recode(agegrp=6,16,@recode(agegrp=7,18,@recode(agegrp=8,22,@recode(agegrp=9,27,@recode(agegrp=10,32,@recode(agegrp=11,37,@recode(agegrp=12,42,@recode(agegrp=13,47,@recode(agegrp=14,52,@recode(agegrp=15,57,@recode(agegrp=16,62,age)))))))))))
  - Series age2=age*age
- Add age and age2 to regression equation

Regression with More Regressors

- Dependent Variable: WAGES
- Method: Least Squares
- Date: 10/19/10 Time: 10:29
- Sample: 1 56529 IF WAGES<8000000 AND WKSWRK<99 AND AGEGRP>5 AND AGEGRP<17
- Included observations: 32765

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-46106.19	2375.629	-19.40800	0.0000
WKSWRK	745.2725	17.73693	42.01811	0.0000
AGE	2125.442	134.6076	15.78991	0.0000
AGE2	-20.15367	1.683832	-11.96893	0.0000

R-squared	0.103843	Mean dependent var	34029.35
Adjusted R-squared	0.103761	S.D. dependent var	49743.57
S.E. of regression	47092.19	Akaike info criterion	24.35772
Sum squared resid	7.27E+13	Schwarz criterion	24.35875
Log likelihood	-399036.4	Hannan-Quinn criter.	24.35805
F-statistic	1265.405	Durbin-Watson stat	2.110714
Prob(F-statistic)	0.000000		

New Regressors

- Age
  - What is the marginal effect of age on the conditional expectation of earnings?

$$\frac{\partial E[Y]}{\partial age} = \beta_{age} + 2\beta_{age2} * age = \$2125 - \$20 * age$$

  - The marginal effect depends on age.
- Why did the coefficient on weeks worked change?
  - Age and age^2 were previously in the error term.
  - They are correlated with weeks worked.