ECON 21110
Applied Microeconometrics
Winter 2022
Lecture 2
Linear Regression Analysis

Eyðfríð Juanna Schrøter Joensen

University of Chicago

January 13, 2022

Linear Regression Analysis

- Linear regression allows us to estimate, and make inferences about unknown population slope coefficients
- The population can be defined as individuals, households, firms, industries, regions, or countries in a given time period
- Ultimately our aim is to estimate the causal effect on Y of a unit change in X

The Multiple Regression Model

- Let $(Y, X_1, ..., X_k, U)$ be a random vector and $(\beta_0, \beta_1, ..., \beta_k)$ be a parameter vector such that:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + U \tag{3.6}$$

- Define $X = (1, X_1, ..., X_k)'$ and $\beta = (\beta_0, \beta_1, ..., \beta_k)'$ such that we can write the model more compactly:

$$Y = X'\beta + U$$

- k independent variables
- β_k reflects the effect of a one unit change in X_k holding all other independent variables fixed
- Key assumption: $\mathbb{E}[U|X_1, X_2, ..., X_k] = \mathbb{E}[U|X] = 0$

The Multiple Regression Model

- Let $(Y_1, \boldsymbol{X}_1), (Y_2, \boldsymbol{X}_2), \ldots, (Y_n, \boldsymbol{X}_n)$ be an i.i.d. sample from (Y, X), where $X = (1, X_1, ..., X_k)'$
- Given the data generating process (3.6) $Y = X'\beta + U$, this means that

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik} + U_i$$

for each sample observation $i = 1, ..., n$. We can write these n equations using short-hand notation:

$$\boldsymbol{Y} = \mathbb{X}\beta + \boldsymbol{U}$$

where the n rows of $\mathbb{X}$ refer to the i.i.d. draws from the population and the $k+1$ columns of $\mathbb{X}$ refer to the independent variables; i.e.

$$\mathbb{X} = \begin{bmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \\ \vdots \\ \boldsymbol{X}_n' \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \ldots & X_{1k} \\ 1 & X_{21} & \ldots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \ldots & X_{nk} \end{bmatrix}, \ \boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \ \boldsymbol{U} = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix}$$

OLS Estimator

- The Ordinary Least Squares (OLS) estimator is given by:

$$\begin{aligned}\widehat{\beta} &= (\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i')^{-1}(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i Y_i) \\ &= (\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i')^{-1}(\sum_{i=1}^{n}\boldsymbol{X}_i Y_i)\end{aligned}$$

  or $\widehat{\beta} = (\mathbb{X}'\,\mathbb{X})^{-1}\,\mathbb{X}'\,\boldsymbol{Y}$ using short-hand notation

- Derived from (i) the Method of Moments (MM) – starting from the moment condition $\mathbb{E}[XU] = 0$ implied by MLR.4 – or (ii) as the best linear approximation to the conditional expectation function $\mathbb{E}\,[Y|X]$. Both these derivations identify $\beta = \mathbb{E}[X'X]^{-1}\,\mathbb{E}[XY]$ in the population model

MLR Assumptions for Unbiasedness of OLS Estimator

- MLR.1: Linear in Parameters, i.e. $Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + U$
- MLR.2: Random Sampling
- MLR.3: No perfect Collinearity
  - No independent variable is a constant, and there are no exact linear relationships among the independent variables
  - For example, violated when including year, cohort, and age in the same regression as $year - cohort = age$
- The key assumption is MLR.4: Zero Conditional Mean

$$\mathbb{E}\left[U | X_1, X_2, ..., X_k\right] = 0$$

Unbiasedness of OLS Estimator

- **Theorem 3.1**: Under MLR.1-MLR.4, we get unbiasedness of the OLS estimates
$$\mathbb{E}\left[\widehat{\beta}_j\right] = \beta_j,\ j = 1, ..., k \tag{3.37}$$
for any values of the population parameters β_j

OLS Fitted Values and Residuals

- Fitted values:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + \widehat{\beta}_2 X_{i2} + ... + \widehat{\beta}_k X_{ik} = \boldsymbol{X}_i' \widehat{\beta}$$

- Residuals:

$$\widehat{U}_i = Y_i - \widehat{Y}_i$$

where

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{U}_i = 0 \text{ and } \frac{1}{n}\sum_{i=1}^{n} X_{ij}\widehat{U}_i = 0$$

for all $j \in \{1, ..., k\}$. That is, the columns of $\mathbb{X}$ are orthogonal to $\widehat{\boldsymbol{U}}$

$$\mathbb{X}'\widehat{\boldsymbol{U}} = \begin{bmatrix} \sum_{i=1}^{n} \widehat{U}_i \\ \sum_{i=1}^{n} X_{i1}\widehat{U}_i \\ \vdots \\ \sum_{i=1}^{n} X_{ik}\widehat{U}_i \end{bmatrix} = 0$$

The Multiple Regression Model: Intuition

- To illustrate what multiple regression does, consider the equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + U$$

- Now, suppose we estimate the regression:

$$X_1 = \delta_1 X_2 + ... + \delta_{k-1} X_k + \epsilon_1$$

keep the residuals, $\widehat{\epsilon_1}$, and then regress Y on them, i.e.

$$Y = \alpha_0 + \alpha_1 \widehat{\epsilon_1} + V$$

Then $\widehat{\beta_1} = \widehat{\alpha}_1$ (we could repeat this process for any of the other X_j)

- The bottom line is that the estimated effect of a variable X_j reflects the relationship between Y and the unique variation in X_j
- This is often called the "*partialling out*" MLR interpretation

OLS Estimator, Revisited

- The Ordinary Least Squares (OLS) Estimator is given by:

$$\begin{aligned}\widehat{\beta}_j &= \frac{\sum_{i=1}^{n} \widehat{\epsilon}_{ij} Y_i}{\sum_{i=1}^{n} \widehat{\epsilon}_{ij}^2} \\ &= \frac{Cov\left(\widehat{\epsilon}_{ij}, Y_i\right)}{Var\left(\widehat{\epsilon}_{ij}\right)}\end{aligned}$$

where $\widehat{\epsilon}_{ij}$ are the OLS residuals from a linear regression of X_j on all the other independent variables $X_1, ..., X_{j-1}, X_{j+1}, ..., X_k$

Unbiasedness of OLS Estimator, Revisited

- MLR.4 is very strong as it implies that the error term is uncorrelated with *all* combinations of the independent variables
- This is the assumption we need for *all* the estimated parameters to be unbiased
- However, this is typically not our goal. In practice, we are most often only concerned with one particular parameter
    - For example, if we are interested in estimating the returns to schooling, we may not be concerned about getting a biased estimate for IQ
- Fortunately, the condition for getting an unbiased estimate of a single parameter is weaker:
  Let ϵ_j denote the error term in the regression of X_j on all the other independent variables. Then $\widehat{\beta}_j$ is unbiased under the weaker mean independence assumption, MLR.4': $\mathbb{E}\left[U|\epsilon_j\right] = 0$ and consistent if ϵ_j is uncorrelated with U

A Look at the Data, SLR

- For example, suppose we estimate the model:

$$log(wage) = \beta_0 + \beta_1 educ + U \tag{S1}$$

  and find that every additional year of education predicts a 5% higher wage
- Does this reflect a causal effect of education on wages? Probably not!
- The reason is that SLR.4, $\mathbb{E}[U|educ] = 0$, is unlikely to hold
- In particular, we expect people with high ability to be more likely to acquire more formal education
- Note that this problem is NOT solved by increasing sample size, n
- In other words, OLS is biased; i.e. $\mathbb{E}\left[\widehat{\beta}_1\right] \neq \beta_1$

A Look at the Data, MLR

- One way to come closer to estimating a causal effect is to include more explanatory variables. In the case of education, we could estimate the model:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + V \tag{M1}$$

  where we let V denote the error term to mark the difference compared to U in (S1)

- This is an example of a multiple regression
- To see the difference between (S1) and (M1), note that $U = \beta_2 IQ + V$
- Equation (M1) takes out IQ from the error term

A Look at the Data, MLR II

- In equation (M1), $\widehat{\beta}_1$ captures how wage changes when *education* increases and *IQ* is held constant
- We say that we "control for IQ" or "hold IQ fixed"
- Equation (S1) and (M1) answer different questions:
  - In equation (S1), we compare the wages of workers with different levels of *education*
  - In equation (M1), we compare the wage of workers who have different levels of *education*, but the same *IQ* score
- The key assumption in interpreting $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in (M1) as reflecting causal effects is that $\mathbb{E}\left[V|educ, IQ\right] = 0$
  - In other words, we assume that the unobservable factors that affect wage are the same for all combinations of *education* and *IQ*

A Look at the Data, MLR III

Multiple Regression

- **Multiple linear regression** (MLR) requires us to think about the sources of variation in the dependent variable
- For example:
  - Differences in wages among *college graduates* with the same degree is not explained by differences in educational attainment
  - Differences in wages between *college graduates* and the *general population* in the US is likely to be partly explained by educational attainment
  - Thus, the variation in wages among college graduates will be used to identify the effect of IQ, but *not* the effect of education
  - In contrast, wage differences between college graduates and (equally clever) people without higher education is used to identify the effect of education

Comparing Simple and Multiple Regression

- Consider a simple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + U$$

and the multiple linear regression:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$$

Let $\widetilde{\beta}_1$ denote the estimate from the SLR and $\widehat{\beta}_1$ and $\widehat{\beta}_2$ the estimates from the MLR. It turns out that:

$$\widetilde{\beta}_1 = \widehat{\beta}_1 + \widehat{\beta}_2 \widetilde{\delta}_1 \tag{3.23}$$

where and $\widetilde{\delta}_1$ is the coefficient estimate from the regression:

$$X_2 = \delta_0 + \delta_1 X_1 + U$$

We thus have that $\widetilde{\beta}_1 = \widehat{\beta}_1$ if

1. $\widehat{\beta}_2 = 0$ or
2. $\widetilde{\delta}_1 = 0$

Omitted Variable Bias I

- Suppose we estimate the model

$$Y_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{i1} + \widehat{U}_i$$

  But the true model is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + U_i$$

- For example, Y_i could be the hourly wage rate, X_{i1} years of schooling, and X_{i2} ability
- Does leaving out X_{i2} from the estimated model affect $\widehat{\beta}_1$?

Omitted Variable Bias II

- Using (3.23), we get:

$$\begin{aligned} \mathbb{E}\left[\widehat{\beta}_1\right] &= \beta_1 + \beta_2\widetilde{\delta}_1 \\ \mathbb{E}\left[\widehat{\beta}_1\right] - \beta_1 &= \beta_2\widetilde{\delta}_1 \end{aligned}$$

  where

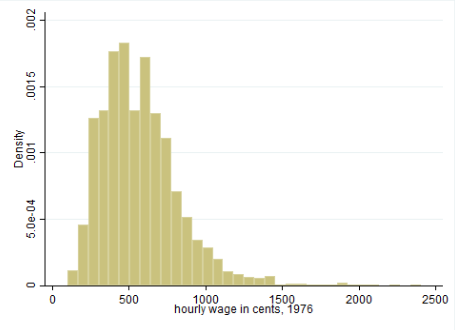
$$\widetilde{\delta}_1 = \frac{Cov(X_{1i}, X_{2i})}{Var(X_{1i})}$$

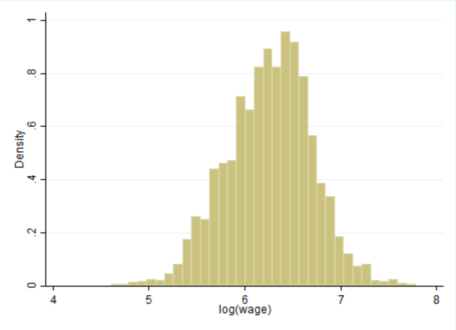
  which has the same sign as the (sample) correlation between X_{1i} and X_{2i}

Omitted Variable Bias III

	$Corr(X_1, X_2) \geq 0$	$Corr(X_1, X_2) \leq 0$
$\beta_2 \geq 0$	Positive Bias	Negative Bias
$\beta_2 \leq 0$	Negative Bias	Positive Bias

A Look at the Data: Variation in Dependent Variable





A Look at the Data I

- We estimate regression (S1) and (M1) above using a sample of 3,010 young men from the NLS data used in Card (1995) with information on monthly wages (*wage*), years of schooling (*educ*) and a measure of cognitive ability (*IQ*)
  - The measure of cognitive ability has been normalized to have a mean of 0 and variance of 1

- Estimating the simple regression of log wages on education:

$$log(wage) = \beta_0 + \beta_1 educ + U \tag{S1}$$

we get $\widetilde{\beta}_1 = 0.052$, implying that one additional year of schooling predicts an (approximately) 5.2% higher wage log(Y)

A Look at the Data II

- We then estimate the multiple regression

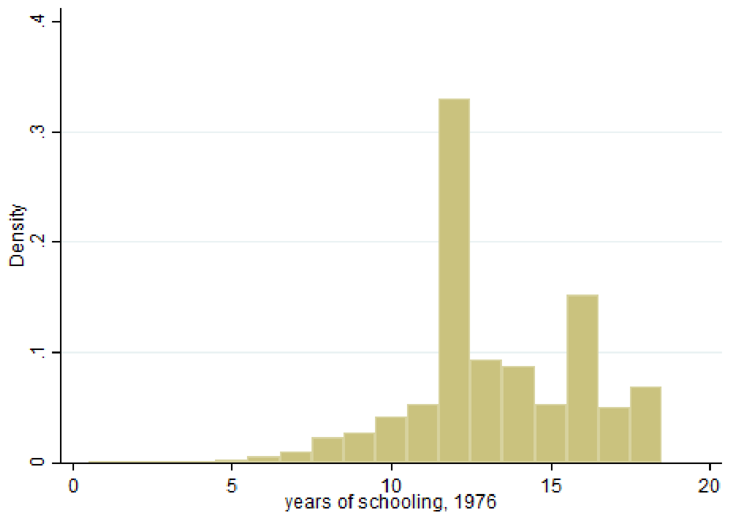
$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + V \tag{M1}$$

  and get $\widehat{\beta}_1 = 0.021$ and $\widehat{\beta}_2 = 0.162$, implying that one additional year of schooling predicts a 2.1% higher wage while an increase in cognitive ability by one standard deviation predicts a 17.6% higher wage

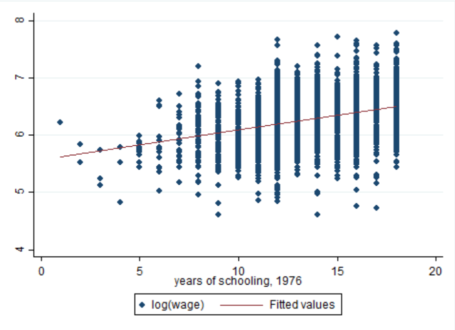
  std. coeff.

- Adding cognitive skill to the regression thus cuts the coefficient for education from 0.052 to 0.021, or almost 60%

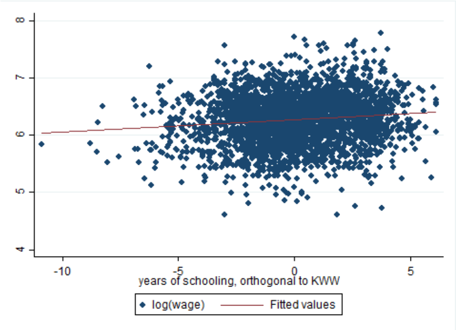
A Look at the Data: Variation in Explanatory Variable



A Look at the Data: MLR.3 (Independent) Variation in Explanatory Variable



Variation in $X_j = educ$



Variation in $\widehat{\epsilon}_j = educ - \widehat{\alpha}_1 KWW$

A Look at the Data III

- Adding cognitive ability to the regression thus cuts the coefficient for education from 0.052 to 0.021, or almost 60%
- To see why, we estimate the regression:

$$IQ = \delta_0 + \delta_1 educ + U$$

and get $\widetilde{\delta}_1 = 0.183$, implying that one more year of schooling predicts a 0.183 standard deviations higher score on cognitive ability

- Now, let's compute the right-hand side of equation (3.23)

$$\begin{aligned} \widehat{\beta}_1 + \widehat{\beta}_2 \widetilde{\delta}_1 &= 0.021 + 0.162 * 0.183 \\ &\approx 0.052 \\ &= \widetilde{\beta}_1 \end{aligned}$$

In other words, the implied OVB is around 0.03

A Look at the Data IV

- Essentially, equation (3.23) splits the correlation between wages and education $\left(\widetilde{\beta}_1\right)$ into two parts:
  - $\widehat{\beta}_1$ represents the part of the correlation between wages and education that holds for given levels of cognitive skills
  - $\widehat{\beta}_2\widetilde{\delta}_1$ represents the part of the correlation which is due to men with higher level of education having higher cognitive skills, and men with higher cognitive skills having higher wages

Omitted Variable Bias: A Look at the Data

- If we are willing to assume that equation (M1) above represents the true population model (a big "IF" as things other than education and cognitive skills are likely to affect wages), then the difference between $\widetilde{\beta}_1$ and $\widehat{\beta}_1$ (0.030) equals the omitted variable bias

- Of course, we never know what the exact true model looks like, but the fact that the estimated returns to education is sensitive to controlling for cognitive skill strongly suggests that the simple regression (S1) suffers from an omitted variable bias problem, even though it is hard to know exactly how strong this bias is

Omitted Variable Bias with more than Two Variables

- Things get more complicated when there are more than one variable in the estimated model
- Suppose the true model has three variables (X_1, X_2, X_3) but only X_1 and X_2 are included in the estimated model
- Then β_2 could be biased even if X_2 is uncorrelated with X_3 as long as X_1 is correlated with X_2 and X_3
  - The relevant question is not whether X_2 is correlated with X_3, but whether ϵ_2 and ϵ_3 from the regressions $X_2 = \delta_0 + \delta_1 X_1 + \epsilon_2$ and $X_3 = \delta_0 + \delta_1 X_1 + \epsilon_3$ are uncorrelated
- However, assuming that X_2 is uncorrelated with both X_1 and X_3, we could get an *approximation* to the omitted variable bias of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ along the lines of the two-variable case above

Functional Form: Higher Order Terms

- Multiple regression is also useful for better capturing functional relationships between variables
- For example, the relationship between wages and labor market experience is typically strictly concave (i.e. the return to experience is decreasing over time)
- One way to capture this is to include both a linear and a squared term for experience:

$$log(wage) = \beta_0 + \beta_3 exp + \beta_4 exp^2 + U \tag{S2}$$

- Note that "ceteris paribus" comparisons are meaningless in this case – it is not possible to hold *experience* constant while changing *experience*^{2}
- $\beta_3 > 0$ and $\beta_4 < 0$ implies that wages are strictly concave in *experience*

Functional Form: A Look at the Data I

- Does the return to experience decrease with years of experience?

- To find out, we extend equation (M1) from above with a quadratic in experience (*exp*)

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 exp + \beta_4 exp^2 + U \qquad \text{(M2)}$$

- Estimating (S2) and (M2) using the NLS sample of young men, we cannot reject a non-linear relationship; i.e. we reject $H_0 : \beta_4 = 0$

Functional Form: A Look at the Data II

- The estimates from (S2) indicate that the return to experience is decreasing in experience since

$$\begin{aligned} \frac{\partial \widehat{log(wage)}}{\partial exp} &= \widehat{\beta}_3 + 2\widehat{\beta}_4 exp \\ &= 0.050 - 0.004exp \end{aligned}$$

- The return to one more year of experience for someone with 10 years of experience is thus

$$0.050 - 2 * 0.002 * 10 = 0.010$$

while the return to someone with only one year of experience is 0.046

- Should we control for experience? or age? or both?
What about MLR.3?

Functional Form: A Look at the Data III

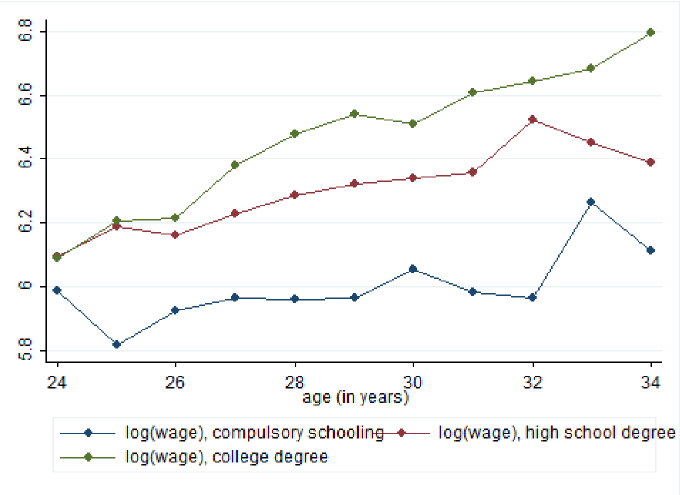
- The estimates from (M2) indicate that the return to experience is decreasing in experience since

$$
\begin{aligned}
\frac{\partial \widehat{log(wage)}}{\partial exp} &= \widehat{\beta}_3 + 2\widehat{\beta}_4 exp \\
&= 0.066 - 0.004 exp
\end{aligned}
$$

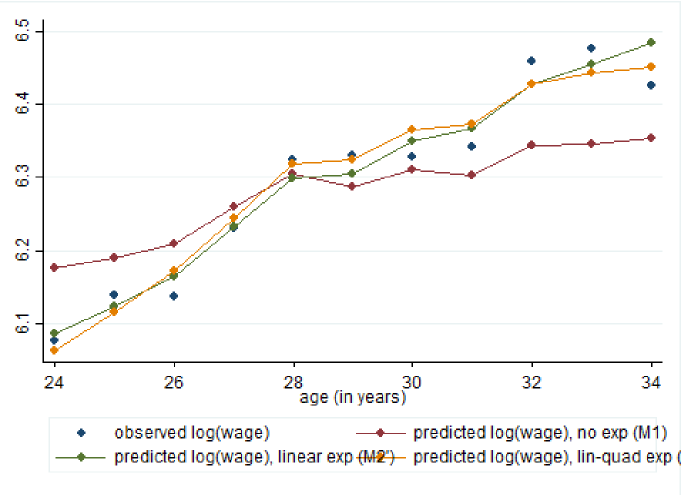
- The return to one more year of experience for someone with 10 years of experience is thus 0.026, while the return to someone with only one year of experience is 0.062. Thus an additional year of experience increases wages by $\approx 6.5\%$ for someone with only one year of experience and by $\approx 2.6\%$ for someone with 10 years of experience

- Should we control for experience? Potential vs actual experience

A Look at the Data: Wages, Education, and Age



A Look at the Data: Assessing Fit of Wage-Age Profiles



Functional Form: Experience and Education

- Say we settle on controlling for potential experience as in (M2), but for the sake of illustration we go back to (M1)
- How should education enter the regression model?
  - Does the return to education increase or decrease with years of education?
  - Should we allow for degree premiums?
  - Does the return to education increase or decrease with IQ?
  - Does the return to experience increase or decrease with years of education?
  - ...

Functional Form: Education
Interaction Effects

- Does the wage return to education vary with IQ?
- To answer this question, we estimate the linear regression:

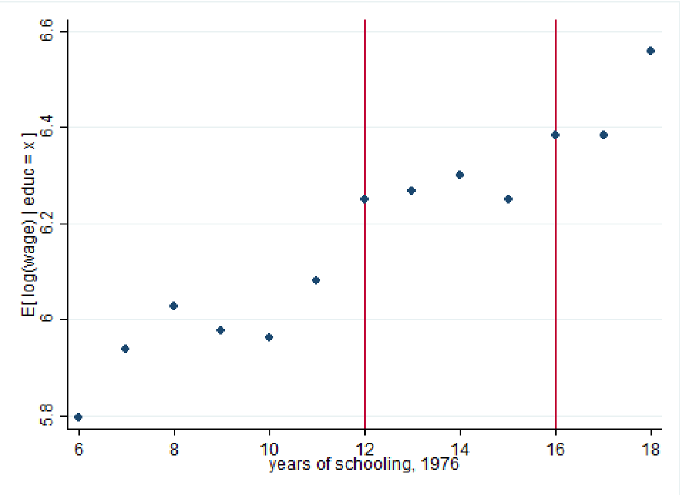
$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 educ * IQ + V$$

- If $\beta_3 > 0$, then the return to education is higher for more intelligent individuals. However, in our example β_3 is not statistically different from zero
- We must take care to interpret the parameters to the original variables β_1 and β_2 correctly
- In this case, we cannot reject $H_0 : \beta_3 = 0$

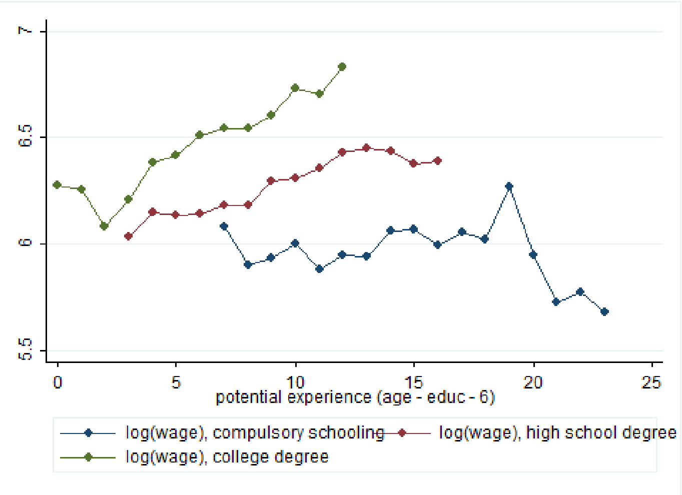
Functional Form: Education

- Return to education does not seem to vary significantly with IQ. Thus highly intelligent men do not seem to benefit more from an additional year of education than less intelligent men
- Return to education does not seem to increase or decrease with years of education either, since $educ^2$ is not statistically significant when added to (M1)
- We still need to determine:
  - whether we should allow for degree premiums?
  - whether the return to experience increase or decrease with years of education?

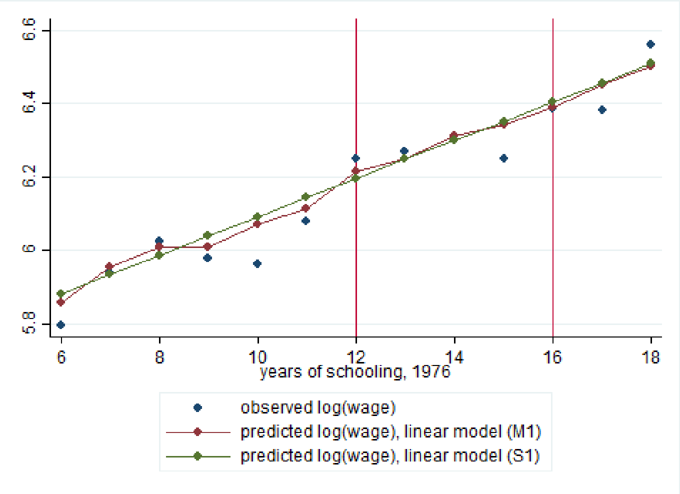
A Look at the Data:
Relationship between Dependent and Explanatory Variable



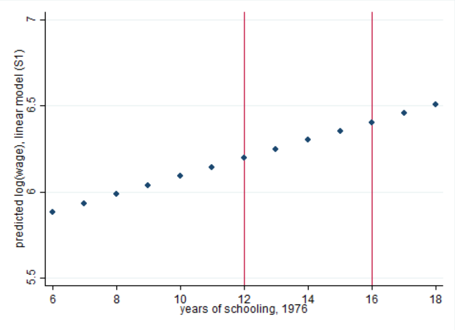
A Look at the Data:
Wages, Education, and Potential Experience



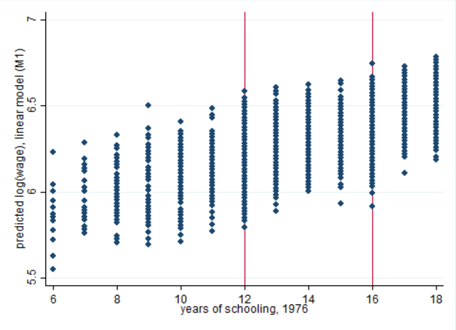
A Look at the Data: Assessing Fit of Linear (in *educ*) Specification



A Look at the Data: Fitted Values from Linear (in *educ*) Specification

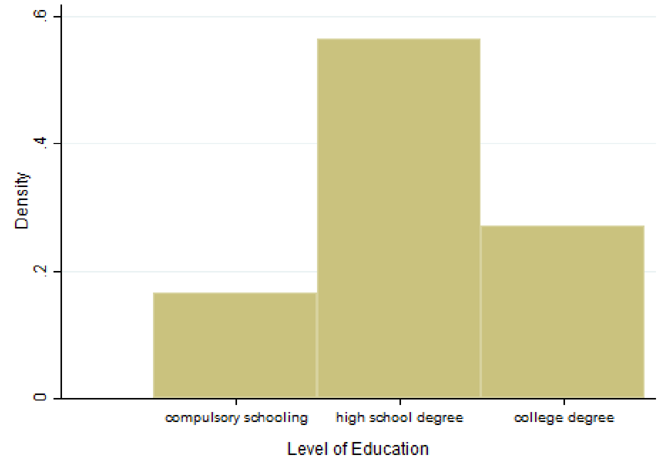


$\widehat{log(wage)}$ from (S1)



$\widehat{log(wage)}$ from (M1)

A Look at the Data: Variation in Explanatory Variable (Degrees)



A Look at the Data: Degree Premiums

- Should we allow for degree premiums?
- Estimating the simple regression of log wages on education:

$$log(wage) = \beta_0 + \beta_1 HS + \beta_2 College + U \tag{SD1}$$

where $HS = 1\,[12 \leq educ < 16]$ and $college = 1\,[educ \geq 16]$

- We get $\widetilde{\beta}_1 = 0.267$ and $\widetilde{\beta}_2 = 0.435$ ▸ Qualitative Information
- College graduates have significantly higher wages than high school graduates; i.e. we reject $H_0 : \beta_1 = \beta_2$
- β_0 denotes the average log(wage) for those without a high school degree; i.e. the reference group

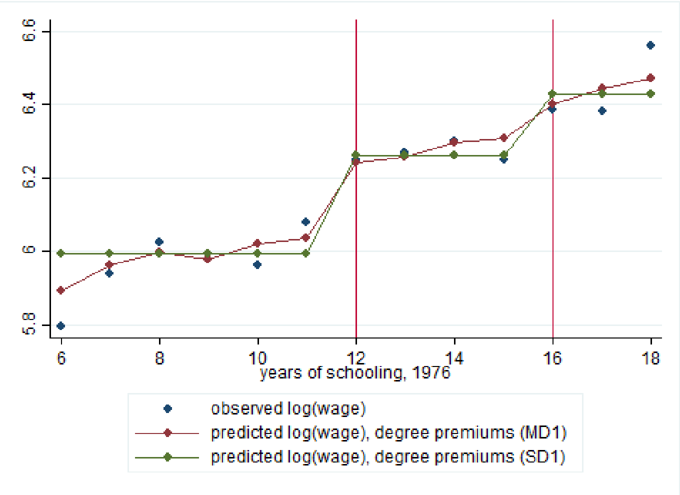
A Look at the Data: Degree Premiums

- Should we allow for degree premiums?
- Estimating the multiple regression of log wages on education:

$$log(wage) = \beta_0 + \beta_1 HS + \beta_2 College + \beta_3 IQ + U \tag{MD1}$$

- We get $\widetilde{\beta}_1 = 0.125$ and $\widetilde{\beta}_2 = 0.195$, implying that getting a high school degree implies $\approx 12.5\%$ higher wage while getting a college degree implies getting $\approx 19.5\%$ higher wage than those without a high school degree
- College graduates *only* seem to have higher wages than high school graduates because they are more intelligent; i.e. we cannot reject $H_0 : \beta_1 = \beta_2$

A Look at the Data:
Assessing Fit of Degree Premium Specification



Functional Form: Education

- Seems like we should allow for degree premiums
  - ...but does the return to experience increase or decrease with years of education?
- We will know after solving Assignment 1

Goodness-of-Fit: R^2

- Like in the simple regression case:

$$R^2 \equiv \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

- R^2 gives the fraction of the sample variation in Y that is explained by *all* the explanatory variables, $X_1, X_2, ..., X_k$
- R^2 *always* increases when more variables are added to the regression
- Low R^2 implies that it is difficult to predict individual outcomes

Goodness-of-Fit: Adjusted $\overline{R}^2$

- Most prefer "adjusted" $\overline{R}^2$

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{SST}$$

- "Adjusted" $\overline{R}^2$ may increase or decrease with the addition of another regressor. Even though SSR still decreases, $\frac{n-1}{n-k-1}$ increases. Which of these two effects dominates is indeterminate
- $\overline{R}^2 \leq R^2$, since $\frac{n-1}{n-k-1} \geq 1$
- Hence, $\bar{R}^2 \leq 1$. On the other hand, unlike R^2, $\bar{R}^2$ may be negative

Goodness-of-Fit: Adjusted $\overline{R}^2$ II

- The adjusted $\overline{R}^2$ is useful for determining the proper functional form of independent variables
- If adjusted $\overline{R}^2$ does not improve from adding higher order terms, then it is better to stick with a simple functional form
- R^2 and adjusted $\overline{R}^2$ also tell us whether there are many other factors in addition to the variables at hand that explain variation in the data
- For example: The R^2 of a regression of log(infant mortality) on log(GDP per capita) is about 0.90
- Question: Why is it appropriate to use logs in this case?

Goodness-of-Fit: A Look at the Data

- How much of the variation in wages can be explained by differences in educational attainment?
- The R^2 from the simple regression of wages on education (S1) is 0.099
- What if we add cognitive ability? The R^2 from regression (M1) is 0.196
- What about functional form? The R^2 from regression (M2) with linear and quadratic terms in experience is 0.234, while the R^2 from regression (M2') with only a linear term in experience is 0.225
- In this case, adding additional variables is far more important for explaining the variation in the data than fine tuning functional form

Model Overview: Comparing Linear Specifications

	(S1)	(M1)	(M2')	(M2)
educ	0.0521***	0.0213***	0.0555***	0.0572***
	(0.0029)	(0.0031)	(0.0045)	(0.0045)
KWW		0.162***	0.121***	0.116***
		(0.0083)	(0.0090)	(0.0090)
exper			0.0266***	0.0656***
			(0.0025)	(0.0072)
expersq				-0.00195***
				(0.0003)
_cons	5.571***	5.981***	5.291***	5.109***
	(0.0388)	(0.0423)	(0.0774)	(0.0831)
N	3010	2963	2963	2963
R^2	0.099	0.196	0.225	0.234
adj. R^2	0.098	0.195	0.225	0.233

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Model Overview: Comparing Degree Premium Specifications

	(SD1)	(MD1)	(MD2')	(MD2)
HS	0.267***	0.125***	0.237***	0.213***
	(0.0215)	(0.0215)	(0.0242)	(0.0244)
college	0.435***	0.195***	0.399***	0.413***
	(0.0240)	(0.0255)	(0.0330)	(0.0329)
KWW		0.161***	0.133***	0.128***
		(0.0082)	(0.0085)	(0.0085)
exper			0.0218***	0.0656***
			(0.0023)	(0.0075)
expersq				-0.00218***
				(0.0004)
_cons	5.993***	6.140***	5.829***	5.659***
	(0.0189)	(0.0193)	(0.0378)	(0.0468)
N	3010	2963	2963	2963
R^2	0.099	0.199	0.223	0.233
adj. R^2	0.098	0.199	0.222	0.232

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Model Selection

- There are no rules set in stone for which variables to include in a model
- In practice, a combination of a priori reasoning and statistical analysis is often used
- Need to think about which factors are correlated with the variable of interest and the dependent variable
  - For example: Educational attainment positively correlated with intelligence
  - Use intelligence test score as a control variable in a regression of wages on education
- Multiple regression models with different sets of variables are typically estimated

Overcontrolling

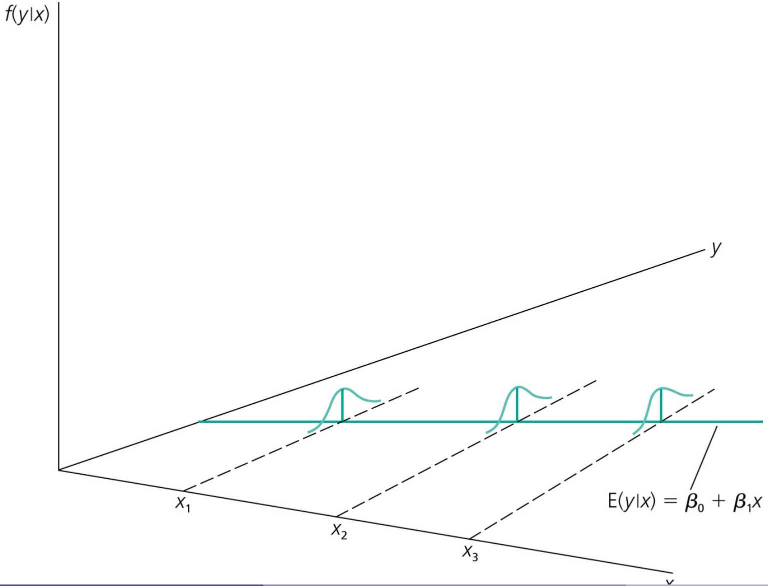
- Suppose you want to estimate the effect of schooling on wages (β_1)
- Would it be right to include a measure of productivity as an independent variable?
- No! We expect schooling to affect the wage because it improves productivity
- Controlling for productivity implies that the schooling coefficient captures the relationship between wages and schooling for a *given level of productivity*

From Bias to Precision...

- Suppose we have specified the linear regression model such that MLR.1-MLR.4 (in particular) hold; i.e. we have an unbiased and consistent estimator
- Should we also consider how precise our estimates are?

Variance of OLS Estimators: Homoskedasticity

Assumption MLR.5: $Var\left(U|X_1,...,X_k\right)=\sigma_U^2$



Variance of OLS Estimators I

- **Theorem 3.2**: Under Assumptions MLR.1-MLR.5

$$Var\left(\widehat{\beta}_j\right) = \frac{\sigma_U^2}{SST_j\left(1 - R_j^2\right)} \text{ for } j = 1, ..., k$$

  where $SST_j = \sum_{i=1}^{n}(X_{ij} - \overline{X_j})^2$ is the total sample variation in X_j and R_j^2 is the R^2 from regressing X_j on all the other independent variables
- We recognize σ_U^2 and SST_j from the simple regression case while R_j^2 is new
- High R_j^2 implies that there is little unique variation in X_j, which means that it is hard to disentangle effect of X_j from the other X's
- Using short-hand notation, $Var\left(\widehat{\beta}\right) = \sigma_U\left(\mathbb{X}'\mathbb{X}\right)^{-1}$

Variance of OLS Estimators II

- As for the SLR, parameters are precisely estimated when:
  - Factors not included in the model are unimportant for the outcome of interest (low σ_U^2)
  - Sample size is large (large n)
  - The variance in the explanatory variable of interest is high (high SST_j)
- But this is not enough in case of MLR:
  The variation in the explanatory variable of interest cannot be too strongly correlated with the other independent variables

Variance of OLS Estimators III

- As for the simple regression case, we need an estimator for σ_U^2 to get an estimate of the variance of parameters
- **Theorem 3.3**: under MLR.1-MLR.5

$$\widehat{\sigma}_U^2 = \frac{\widehat{\boldsymbol{U}}'\widehat{\boldsymbol{U}}}{n-k-1} = \frac{\sum_{i=1}^{n} \widehat{U}_i^2}{n-k-1} = \frac{SSR}{n-k-1}$$

  is an unbiased estimator of σ_U^2

- Note that $n-k-1$ refers to the degrees of freedom
- Adding additional variables to the regression implies that SSR and $n-k-1$ fall, which implies that the net effect on $\widehat{\sigma}_U^2$ is ambiguous

Standard Error of OLS Estimators

- To conduct hypothesis testing, we need to estimate the **standard deviation** of $\widehat{\beta}_j$, for $j = 1, 2, ..., k$

$$\sigma\left(\widehat{\beta}_j\right) = \sqrt{Var\left(\widehat{\beta}_j\right)} = \frac{\sigma_U}{\sqrt{SST_j\left(1 - R_j^2\right)}}$$

- Since σ_U is unknown, we replace it with its estimator, $\widehat{\sigma}_u$, giving us the **standard error** of $\widehat{\beta}_j$

$$\widehat{\sigma}\left(\widehat{\beta}_j\right) = \frac{\widehat{\sigma}_U}{\sqrt{SST_j\left(1 - R_j^2\right)}}$$

- In practice, we focus on the standard error of parameters, not their variances

Efficiency of OLS

- **Theorem 3.4**: Under MLR.1-MLR.5, the OLS estimator $\widehat{\beta}_j$ is the **Best Linear Unbiased Estimator** (**BLUE**) of β_j
- OLS is both *unbiased* (correct on average) and *efficient* (exhibits a low variance)
- This is often called the Gauss-Markov Theorem

Variance of OLS Estimators: A Look at the Data I

- Suppose we are interested in which *type* of cognitive ability matters most for outcomes
- The data from Card (1995) contains two measures of cognitive ability: KWW used in the previous slides and IQ
- First, consider a version of regression (M1) for those with non-missing IQ:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 KWW + U \qquad \text{(M1')}$$

Estimating (M1') by OLS, we get $\widehat{\beta}_2 = 0.149$ and $\widehat{\sigma}(\widehat{\beta}_2) = 0.0107$, implying that the association between KWW and wages is quite precisely estimated. Why?

  - relatively large sample ($2,040$ men with non-missing IQ)
  - Substantial variation in KWW
  - $educ$ and KWW moderately positively correlated ($\rho = 0.38$)

Variance of OLS Estimators: A Look at the Data II

- Now, let's add our measure of IQ to the regression:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 KWW + \beta_3 IQ + V \qquad \text{(M1")}$$

- Estimating (M1") by OLS, we get $\widetilde{\beta}_2 = 0.142$ and $\widehat{\sigma}(\widetilde{\beta}_2) = 0.0112$
- Adding a measure of IQ slightly reduces the estimated effect of *KWW* (why?) *and* slightly decreases the precision of the estimate
  - The reason for the increase in variance is that, since KWW and IQ are positively correlated ($\rho = 0.43$) there is less variation in the data that is unique to KWW
  - Regressing *KWW* on *educ* and *IQ* gives an R^2 of 0.22, implying that $(1 - 0.22) * 100 = 88\%$ of the variation in KWW can be used to estimate $\widetilde{\beta}_2$ in (M1")

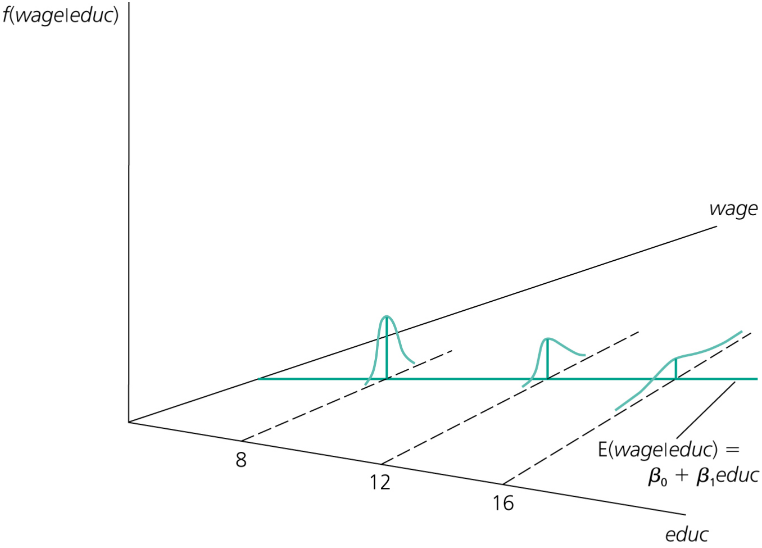
What is Heteroskedasticity?

- **Heteroskedasticity** means that Assumption MLR.5 of homoskedasticity is violated; i.e.

$$Var\left(U|X_1, ..., X_k\right) \neq \sigma_U^2$$

- That is, the error variance is *not* the same across all values of the independent variables

Illustration: Heteroskedasticity



How does Heteroskedasticity matter?

- $\widehat{\beta}_j$ remains unbiased in the presence of heteroskedasticity
- But $Var\left(\widehat{\beta}_j\right)$ is biased
  - standard errors and t-statistics not correct
  - we can no longer do hypothesis testing
- The simplest and most common solution to this problem is to calculate heteroskedasticity-robust standard errors

Non-Constant Variance

- Consider the model:

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

- Assume Assumptions MLR.1-4 hold, but instead of Assumption MLR.5 we allow for heteroskedasticity:

$$Var\left(U_i|X_i\right) = \sigma_i^2$$

where the subscript i denotes that the variance depends on the particular value of X_i

Computing the Variance

- Remember that $Var\left(\widehat{\beta}_1\right)$ is a measure for how much $\widehat{\beta}_1$ will vary across different samples of the same size from the same population
- When $Var\left(\widehat{\beta}_1\right)$ is high, different samples give very different values for $\widehat{\beta}_1$
- The variance formula in case of heteroskedasticity is:

$$Var\left(\widehat{\beta}_1\right) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2\sigma_i^2}{SST_X^2} \tag{8.2}$$

- When $\sigma_i^2 = \sigma_U^2$ for all i, this formula reduces to the usual form: $\frac{\sigma_U^2}{SST_X}$

Intuition I

- Let's have a closer look at (8.2) and see if we can build some intuition
- When is
$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2\sigma_i^2}{SST_X^2} > \frac{\sigma_U^2}{SST_X} \ ?$$
- That is, when will heteroskedasticity lead to a larger variance of the estimated coefficients compared to homoskedasticity?
- Answer: When $(X_i - \overline{X})^2$ is *positively correlated* with σ_i^2
- That is, when observations with values of X far from the mean are more strongly affected by factors other than X (= higher variance of U)
- Essentially, $(X_i - \overline{X})^2$ is a weight attached to each observation i. Observations with large or small X_i (relative to the mean) get higher weight

Intuition II

- The key point is that observations far away from the mean value of X are more informative of the slope coefficient, β_1
  - To see this, note that for observations with similar values of X, differences in U are more important than movement along the slope for differences in Y. This gives us little information from which we could infer the slope, β_1
- When σ_i^2 is large for values far away from the mean of X; i.e. when $(X_i - \overline{X})^2$ is positively correlated with σ_i^2, the most informative observations are subject to a lot of noise. Chance then plays a big role for the value of $\widehat{\beta}_1$ in any given sample
  - In other words, $\widehat{\beta}_1$ will vary a lot across samples due to different values of the error term for the most informative observations
- Not taking heteroskedasticity into account in this case will imply that we put too much faith in $\widehat{\beta}_1$ from a given sample; i.e. we underestimate the variance

Intuition III

- What about the opposite case; i.e. if $(X_i - \overline{X})^2$ is *negatively correlated* with σ_i^2?
- Then the realizations of the error term, U, are smaller for informative observations, implying that $Var\left(\widehat{\beta}_1\right)$ is smaller than under homoskedasticity
  - A priori, it is thus not clear that heteroskedasticity makes $Var\left(\widehat{\beta}_1\right)$ larger or smaller - it depends on the type of heteroskedasticity. In practice, however, heteroskedasticity typically inflates $Var\left(\widehat{\beta}_1\right)$
- This has to do with the structure of the data. Typically, σ_i^2 is monotonically increasing or decreasing in X_i, but the variance at either end of the spectrum is disproportionally larger, dominating the low variance of the error term at the other end
  - For example, if people with low income have a low variance in their savings – which reduces $Var\left(\widehat{\beta}_1\right)$, but this effect is dominated by the very large variance in savings for people with high income

An Estimator for the Variance

- White (1980) showed that

$$\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2 \widehat{U}_i^2}{SST_X^2} \tag{8.3}$$

  is a consistent estimator of $Var\left(\widehat{\beta}_1\right)$, where $\widehat{U}_i$ denotes the residuals from the regression of Y on X

- Since it is consistent, it is valid in large samples

Multivariate Case

- Similarly, for the multivariate case:

$$\widehat{Var\left(\widehat{\beta}_j\right)} = \frac{\sum_{i=1}^{n} \widehat{\epsilon}_{ij}^2 \widehat{U}_i^2}{SSR_j^2} \tag{8.4}$$

  where $\widehat{\epsilon}_{ij}$ denotes the ith residual from regressing X_j on all other independent variables, and SSR_j is the sum of squared residuals from this regression

- The square root of (8.4) is called the **heteroskedasticity-robust standard error** of $\widehat{\beta}_j$

Multivariate Case II

- The heteroskedasticity-robust standard errors:

$$\widehat{\sigma}\left(\widehat{\beta}_j\right) = \sqrt{\frac{\sum_{i=1}^{n} \widehat{\epsilon}_{ij}^2 \widehat{U}_i^2}{SSR_j^2}}$$

go under many different names:

  - Huber-White standard errors
  - Sandwich standard errors
  - Robust standard errors

Robust t-statistic

- Hypothesis testing the same as before, but we use the robust instead of the ordinary standard errors
- Robust standard errors are typically somewhat larger than non-robust standard errors, implying smaller t-statistics
- But robust standard errors can be smaller
- Usually not a big difference

Robust t-statistic in Stata

- Type `robust` after the regression equation
- For example: `reg lwage educ exper, robust`
- F- and t-tests following the estimation will also be using the robust standard errors

In Practice...

- In applied work, it is common to use the `robust` standard errors as a default option
- This is ok in case of "large" samples
- We can do better if we know the exact form of the error variance and use Weighted Least Squares (WLS) – more on this when we discuss applications

Testing for Heteroskedasticity: Breuch-Pagan (BP)

- There are several ways to test for hetereoskedasticity
- The **Breusch-Pagan test** proceeds in four steps:
  1. Estimate your model with OLS. Obtain the squared residuals, $\widehat{U}^2$, for each observation
  2. Regress $\widehat{U}^2$ on the independent variables, X's. Keep the R^2 from this regression, $R^2_{\widehat{U}^2}$
  3. Use R to compute the F-statistic

$$F = \frac{R^2_{\widehat{U}^2}/k}{\left(1 - R^2_{\widehat{U}^2}\right)/(n-k-1)} \tag{8.15}$$

  4. Compute the p-value at which the null hypothesis of homoskedasticity can be rejected. A low p-value is a strong indication of heteroskedasticity

Testing for Heteroskedasticity: The White Test

- Regress $\widehat{U}^2$ on all the independent variables, the squares of the independent variables, and the cross-products between the independent variables (8.19)
- The **White test** is the LM test that all of the coefficients in this equation are zero (we could also use an F test)
- An alternative to the White test is to regress $\widehat{U}^2$ on $\widehat{Y}$ and $\widehat{Y}^2$

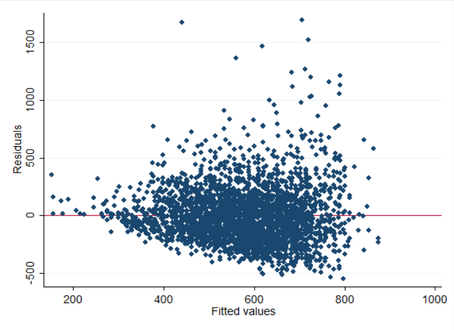
What if we Reject Homoskedasticity?

- Need to undertake some corrective measure
- Alternatives:
  1. Transform the model, for example by using logs, and do the BP/White test again
  2. Use the robust standard errors
  3. Specify the form of the error variance

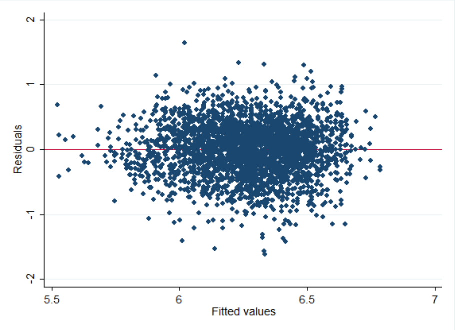
A Look at the Data: Heteroskedasticity

- Why did we estimate (M1) in logs – rather than in levels; i.e. why $log(wage)$ rather than *wage* as the dependent variable?
- A quick look at the data reveals that the standard deviation of *wage* increases with the level of education, but this is not true for $log(wage)$
- If we estimate (M1) in levels – rather than logs – we get $\widehat{\beta}_1 = 12.79$ meaning that an additional year of education increases hourly wages by 12.79 cents on average, where $\widehat{\sigma}\left(\widehat{\beta}_1\right) = 1.88$
- However, both the BP and the White test *reject* the null hypothesis of homoskedasticity; i.e. $H_0 : Var\left(U|X\right) = \sigma_U^2$
- Estimating (M1) in levels with robust standard errors, we get $\widehat{\sigma}\left(\widehat{\beta}_1\right) = 1.89$ – only slight increase, unaffected inference
- Going back to estimating our original (M1) in logs, we *cannot reject* the null hypothesis of homoskedasticity

A Look at the Data: Heteroskedasticity
Graphical Illustration



$\left(\widehat{U}, \widehat{wage}\right)$ from (M1) in levels



$\left(\widehat{U}, \widehat{log(wage)}\right)$ from (M1)

Summary

- Multiple regression analysis makes it possible to control for many factors
- Allows us to come closer to the "gold standard" of a randomized trial IF we observe all the relevant control variables – economic theory and institutional knowledge are our best guides
- Thinking about omitted variables and potential violations of MLR.4 is paramount
- OLS is unbiased under assumptions MLR.1-MLR.4
- OLS is unbiased *and* efficient under assumptions MLR.1-MLR.5

APPENDIX: Supplementary material

Appendix: A Look at the Data, MLR III

- Consider a data set of six workers

ID	*wage*	*educ*	*IQ*
1	20	12	28
2	27	16	28
3	22	12	40
4	33	16	40
5	30	16	40
6	27	16	28

- $\widetilde{\beta}_1$ in (S1) is estimated based on comparing the wages of 1 & 3 vs. 2, 4, 5 & 6
- $\widehat{\beta}_1$ in (M1) is estimated based on comparing the wages of 1 vs. 2 & 6, and 3 vs. 4 & 5
- $\widehat{\beta}_2$ in (M1) is estimated based on comparing the wages of 1 vs. 3, and 2 & 6 vs. 4 & 5

Back: A Look at the Data, MLR II

Appendix: Functional Form I
Logs

- Logs are used frequently in applied research for a number of reasons:

  1. $log(Y)$ often fits the CLM assumptions better (normal errors; homoskedasticity) when $Y > 0$
  2. Often (but not always!) imply a more plausible functional form in linear regressions
     - Example: City size, wages,...
  3. Easy to interpret - approximates percent changes

Appendix: Functional Form II
Logs and Percentage Changes

- Consider the estimated model:

$$\widehat{log(Y)} = \widehat{\beta}_0 + \widehat{\beta}_1 X$$

- The exact percentage change in $\widehat{Y}$ is then:

$$\%\Delta\widehat{Y} = 100\left[e^{\widehat{\beta}_1 \Delta X} - 1\right]$$

- Note that the percentage change depends on whether we increase or decrease X
  - Suppose $\widehat{\beta}_1 = 0.1$. Then $\Delta X = 1$ gives $e^{0.1*1} - 1 \approx 0.105$ while $\Delta X = -1$ gives $e^{-0.1*1} - 1 \approx 0.095$
- Thus $100\widehat{\beta}_1$ is in between the percentage change in $\widehat{Y}$ for increasing or decreasing X

Appendix: Functional Form III
A Frequent Mistake: Logs close to zero

- Suppose we run a regression of annual earnings on a set explanatory variables
- Should we take the log of annual earnings?
- A teenager gets his/her first summer job:
  $log(10,000) - log(1) = 9.2103$
- You get promoted to CEO:
  $log(1,500,000) - log(800,000) = 0.6286$
- Most of the variation in the data will be due to whether someone has a job or not
- The estimated β's capture the effect of the explanatory variables of being in the labor market, not the effect on wages
- Conclusion: Look at the distribution of your dependent variables!

Back

Appendix: Beta Coefficients I

- Beta coefficients:
$$\widehat{b}_j = \left(\frac{\widehat{\sigma}_j}{\widehat{\sigma}_Y}\right)\widehat{\beta}_j$$
- If X_j changes by one standard deviation, then Y changes by $\widehat{b}_j$ standard deviations
- Beta coefficients are useful when the dependent or independent variables that have no natural metric
- Example: A regression of wages on a math test score and a measure of beauty
- Also useful when comparing the effect of different variables

Appendix: Beta Coefficients II
Example

- Suppose we regress log(wages) on an index of beauty measured on a 1-10 scale and obtain $\widehat{\beta}_1 = 0.05$
- $\widehat{\beta}_1 = 0.05$ implies that a one-unit increase in the beauty-index is associated with an approximate 5%-increase in wages
- But what does a one-unit increase in this index really mean? A better way to express the effect is to standardize the coefficient
- Suppose the standard deviation of the beauty index ($\widehat{\sigma}_{beauty}$) is 2. Then a one-standard deviation increase in beauty is associated with a $\approx 10\%$ higher wage ($\widehat{\sigma}_{beauty}\widehat{\beta}_{beauty} = 2 * 0.05 = 0.10$)
  - I need to move to index units in order to increase beauty by one standard deviation, and each index unit increase the wage by $\approx 10\%$
  - When the dependent variable has a natural metric (which log wages does), it is often sufficient just to report the **standardized coefficient**

Appendix: Beta Coefficients III
Example

- In order to get the **beta coefficient**, I need to know how many standard deviations in the distribution of log(wages) an increase by 0.1 corresponds to
- Suppose $\widehat{\sigma}_{wages} = 0.3$. Then

$$\widehat{b}_{beauty} = \left(\frac{\widehat{\sigma}_{beauty}}{\widehat{\sigma}_{wages}}\right)\widehat{\beta}_{beauty} = \frac{\widehat{\sigma}_{beauty}\widehat{\beta}_{beauty}}{\widehat{\sigma}_{wages}} = \frac{2 * 0.05}{0.3} = 0.333...$$

- Increasing beauty by one standard deviation is thus associated with a 1/3 standard deviation increase in log(wages)

Back

Appendix: Qualitative Information I
Binary and Categorical Variables

- We use *dummy variables* (or *indicator variables*) that take the value 0 or 1 to represent binary outcomes
    - Example 1: The variable *female* takes the value 1 in case a person is female and 0 in case he is male; i.e. $female = 1\,[i \text{ is female}]$
    - Example 2: The variable *private* takes the value 1 if person works in the private sector and 0 if he/she works in the public sector; i.e. $private = 1\,[i \text{ works in private sector}]$
- Categorical information is often coded as multiple dummy variables, one for each specific value

Appendix: Qualitative Information II
Interpreting Binary Variables I

- Consider the regression:

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + U$$

- δ_0 gives the difference in conditional means between males and females:

$$\delta_0 = \mathbb{E}[wage|female = 1, educ] - \mathbb{E}[wage|female = 0, educ]$$

- In other words, δ_0 gives the difference in intercept between males and females on the regression line

Appendix: Qualitative Information III
Interpreting Binary Variables II

- Note that adding another variable for *male*, i.e.

$$wage = \beta_0 + \delta_0 female + \delta_1 male + \beta_1 educ + U$$

  would ***not*** work since *female* and *male* are perfectly correlated

- We always need to exclude one category

Appendix: Qualitative Information IV
Multiple Categories I

- Suppose we add an interaction between gender and marriage

$$wage = \beta_0 + \delta_0 female + \delta_1 married + \delta_2 female * married + U$$

- The effect of marriage is the same for males and females under H_0: $\delta_2 = 0$ (two-sided t-test)
- Expected wages: Unmarried males (β_0); unmarried females ($\beta_0 + \delta_0$); married males ($\beta_0 + \delta_1$) and married females ($\beta_0 + \delta_0 + \delta_1 + \delta_2$)
- Note that we include three dummy variables to distinguish between four categories
- A model that contains all possible categories is called **saturated**

Appendix: Qualitative Information V
Multiple Categories II

- It is useful and easy to create dummy variables from a categorical variable
- For example: Suppose *educ* is equal to 1 for only compulsory schooling, 2 for high school, etc.
- In Stata: `tabulate` *educ*, `generate`(*educdummy*)

Appendix: Qualitative Information VI
Multiple Categories III

- Dummy variables are useful when we are uncertain about the true functional form
- Instead of including higher order terms (i.e. quadratic, cubic, etc...), we can let dummy variables denote different segments of the independent variable
- Suppose we have a variable for years of education
- We could enter this as a linear variable or adding higher order terms
- Alternatively, we could include *educdummy*
- This is a form of non-parametric regression

Appendix: Qualitative Information VII
Interactions with Dummy Variables

- Is the return to education the same for males and females?

$$wage = \beta_0 + \delta_0 female + \beta_1 educ + \beta_2 female * educ + U$$

- β_1 is the slope coefficient – the return to education – for men
- β_2 is the difference in the returns to education between males and females
- δ_0 is the difference in intercepts – the expected wage differences between a male and a female with $educ = 0$
- The return to education is the same for males and females under H_0: $\beta_2 = 0$ (two-sided t-test)

Appendix: Qualitative Information VIII
Testing for Different Functional Forms across Groups

- Consider the regression model:

$$wage = \beta_0 + \beta_1 educ + \beta_2 exp + U$$

- Suppose we want to test if there is *any* difference in the intercept or slopes of this model between males and females

- We could test this by estimating the model:

$$\begin{aligned} wage =& \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 female * educ \\ &+ \beta_2 exp + \delta_2 female * exp + U \end{aligned}$$

and then do a F-test of the null hypothesis
H_0: $\delta_0 = 0, \delta_1 = 0, \delta_2 = 0$

Appendix: Qualitative Information IX
Chow Test

- A simpler way of testing for difference in functional forms across groups in case of many independent variables is the **Chow test**
- Estimate the model separately for group 1 and 2 to obtain SSR_1 and SSR_2
- Estimate the restricted model on the pooled sample and obtain SSR_P
- Compute the Chow statistic:

$$F = \frac{SSR_P - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \frac{n - 2(k+1)}{k+1}$$

- Reject the null if the statistic is greater than the critical F-value

Back