Vniversitat de València OpenCourseWare

The Simple Regression Model

$$y = \beta_0 + \beta_1 x + u$$

Simple Regression Model

- In writing down a model that "explains *y* in terms of *x*", we must confront three issues:
  1. Since there is never an exact relationship between variables, how do we allow for other factors to affect *y*?
  2. What is the functional relationship between *y* and *x*?
  3. How can we be sure we are capturing a "ceteris paribus" relationship between *y* and *x* (if that is the desired goal)?

Definition

- The Simple Linear Regression (SLR) model resolve these ambiguities by writing down the simple equation

$$y = \beta_0 + \beta_1 x + u$$

- This is a **population model**.

Some Terminology

- In the simple linear regression model, where $y = \beta_0 + \beta_1 x + u$, we typically refer to y as the
  - Dependent Variable, or
  - Left-Hand Side Variable, or
  - Explained Variable, or
  - Regressand

Some Terminology, cont.

- In the simple linear regression of y on x, we typically refer to x as the
  - Independent Variable, or
  - Right-Hand Side Variable, or
  - Explanatory Variable, or
  - Regressor, or
  - Covariate, or
  - Control Variables

Some Terminology, cont.

- In the simple linear regression of *y* on *x*, we typically refer to *u* as the
  - Error Term, or
  - Disturbance

The variable *u* represents factors other than *x* that affect *y*.

Some Terminology, cont.

- In the simple linear regression of *y* on *x*, β_0 and β_1 are unknown parameters of interest to be estimated from a given sample.

Linear Regression Model

- If the other factors in u are held fixed, so that the change in u is zero, $\Delta u = 0$, then x has a linear effect on y:

$$\Delta y = \beta_1 \Delta x \quad \text{if} \quad \Delta u = 0$$

- This means that β_1 is the slope parameter in the relationship between y and x, holding the other factors in u fixed.

Linear Regression Model

- The linearity of $y = \beta_0 + \beta_1 x + u$ implies that a one-unit change in x has the same effect on y, regardless of the initial value of x.

- Given that u is an unobserved random variable, we need an assumption about how x and u are related, otherwise we can't do any progress.

A Simple (non restrictive) Assumption

- The average value of u, the error term, in the population is 0. That is,

$$E(u) = 0$$

- This is not a restrictive assumption, since we can always use β_0 to normalize $E(u)$ to 0. (If you don't believe me try Problem 2.2)

Key Assumption: Zero Conditional Mean

- We need to make a crucial assumption about how u and x are related.
- We want it to be the case that knowing something about x does not give us any information about u, so that they are completely unrelated. That is, that:

  $E(u|x) = E(u) = 0$, which implies

  $E(y|x) = \beta_0 + \beta_1 x$

Key Assumption: Zero Conditional Mean

- Hence the crucial assumption is that the average value of u does not depend on the value of x:

$$E(u|x) = E(u) = 0$$

where the second equality follows from above and is not restrictive as long as we include an intercept, β_0, in the equation.

Key Assumption: Zero Conditional Mean

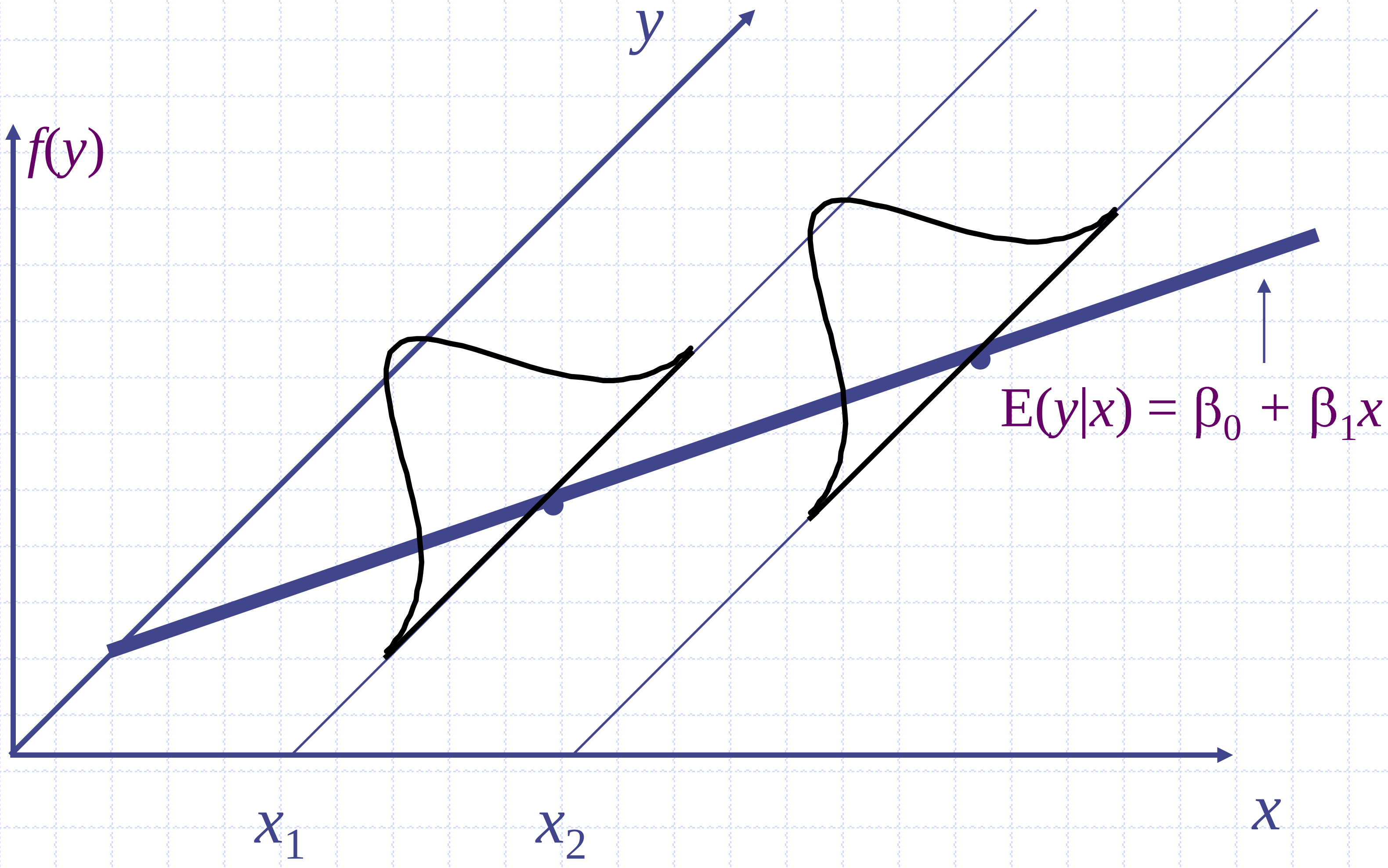
- The equation

$$E(y|x) = \beta_0 + \beta_1 x$$

is known as the **population regression function** (PRF), $E(y|x)$, is a linear function of x.

- The linearity means that a one-unit increase in x changes the *expected value* of y by the amount β_{1}.

E(y|x) as a linear function of x, where for any x the distribution of y is centered about E(y|x)

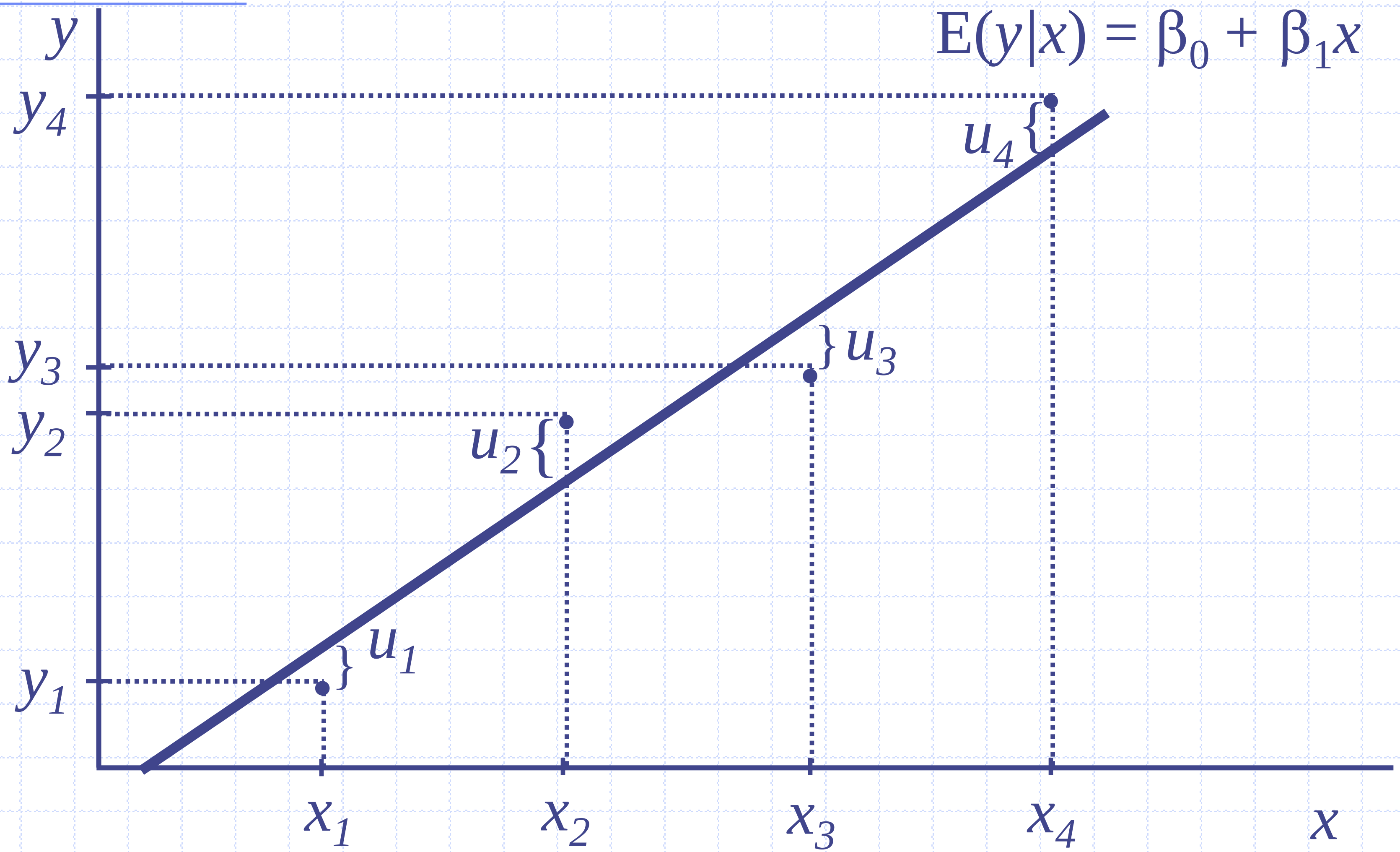


Ordinary Least Squares

- Basic idea of regression is to estimate the population parameters, (β_0 , β_1), from a sample.
- Let $\{(y_i ,x_i): i = 1,\ldots,n\}$ denote a random sample of size n from the population.
- For each observation in this sample, it will be the case that

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Population regression line, sample data points and the associated error terms



Deriving OLS Estimates

- To derive the OLS estimates we need to realize that our main assumption of $E(u|x) = E(u) = 0$ also implies that

$$\mathrm{Cov}(x,u) = E(xu) = 0$$

- Why? Remember from basic probability that $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$

Deriving OLS continued

- These, $E(u) = 0$ and $E(xu) = 0$, are two population restrictions we can write just in terms of x, y, β_0 and β_1 , since $u = y - \beta_0 - \beta_1 x$

$$E(y - \beta_0 - \beta_1 x) = 0$$

$$E[x(y - \beta_0 - \beta_1 x)] = 0$$

- These are called moment restrictions

Deriving OLS using MOM.

- The method of moments approach to estimation implies imposing the population moment restrictions on the sample moments.

- What does this mean? Recall that for E(X), the mean of a population distribution, a sample estimator of E(X) is simply the arithmetic mean of the sample.

More on the Derivation of OLS

- We want to choose values of the parameters that will ensure that the sample versions of our moment restrictions are true.
- The sample versions are as follows:

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

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

More on the Derivation of OLS

- Given the definition of a sample mean, and properties of summation, we can rewrite the first condition as follows

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x},$$

or

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

More on the Derivation of OLS

$$\sum_{i=1}^{n} x_i \left(y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x} \right) - \hat{\beta}_1 x_i \right) = 0$$

$$\sum_{i=1}^{n} x_i \left(y_i - \bar{y} \right) = \hat{\beta}_1 \sum_{i=1}^{n} x_i \left(x_i - \bar{x} \right)$$

$$\sum_{i=1}^{n} \left(x_i - \bar{x} \right) \left(y_i - \bar{y} \right) = \hat{\beta}_1 \sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2$$

So the OLS estimated slope is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

provided that $\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$

Summary of OLS slope estimate

- The slope estimator is the sample covariance between x and y divided by the sample variance of x.
- If x and y are positively correlated, the slope will be positive.
- If x and y are negatively correlated, the slope will be negative.
- We only need x to vary in our sample.

More on the OLS estimates

- Given the OLS estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, the **fitted value** for y when $x = x_i$ is given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- This is the **OLS regression line** or **Sample Regression Function** (SRF). The value that the model predicts for y when $x = x_i$.
- There is a fitted value for each observation in the sample.

More on the OLS estimates

- The SRF is the sample counterpart of the PRF.
- It is important to remember that the PRF, $E(y|x) = \beta_0 + \beta_1 x$, is something fixed, but unknown, in the population. Since the SRF is obtained for a given sample of data, a new sample will generate different estimates.

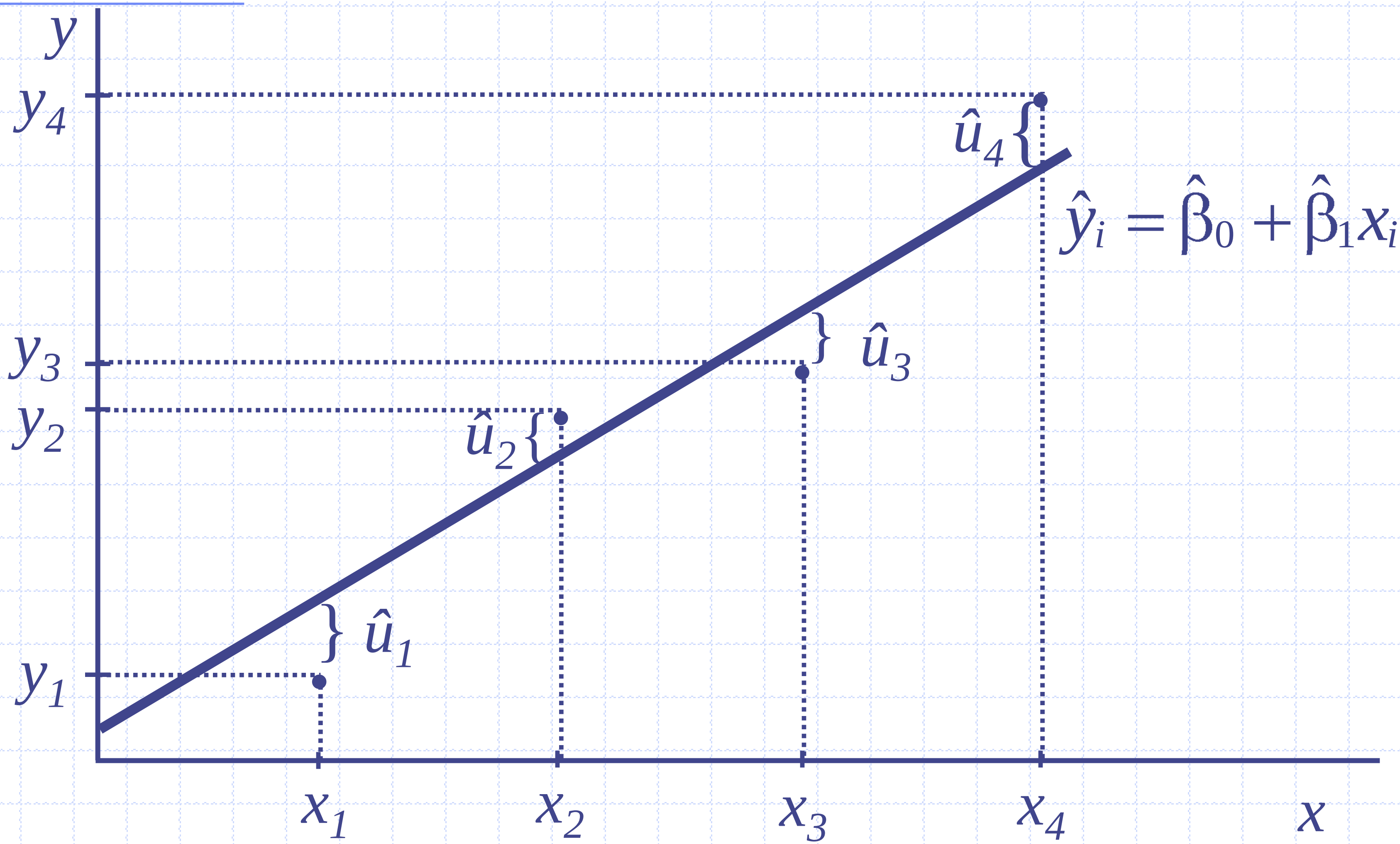
More on the OLS estimates

- The **residual** for observation *i* is the difference between the actual y_i and its fitted value, $\hat{y}_i$,

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

- Again there are *n* residuals.
- The residual, $\hat{u}$, is an estimate of the error term, *u*, and is the difference between the fitted line (SRF) and the sample point.

Sample regression line, sample data points and the associated estimated error terms



More on the OLS estimates

- In most cases the slope estimate, $\hat{\beta}_1$, is of primary interest. This can be written as

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x}$$

- It tell us the amount by which $\hat{y}$ changes when x increases by one unit. Equivalently,

$$\Delta \hat{y} = \hat{\beta}_1 . \Delta x$$

- Given a change in x, we can compute the predicted change in y.

OLS: Alternate approach

- These are called **Ordinary Least Square** (**OLS**) estimates, but we have derived them from two moment conditions. Where does the OLS term come from?
- Intuitively, OLS is fitting a line through the sample points such that the sum of squared residuals is as small as possible, hence the term least squares

OLS: Alternate approach

- Consider b_0 and b_1 two estimators of the population parameters β_0 and β_1 (OLS or any other estimators). The residual for observation *i*, given these estimators, is

$$e_i = y_i - b_0 - b_1 x_i$$

OLS: Alternate approach

- Given the intuitive idea of fitting a line, we can set up a formal minimization problem.
- It can be shown that OLS solves the following optimization problem:

$$\min_{b_0, b_1} \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_i \right)^2$$

OLS: Alternate approach

- If one uses calculus to solve the minimization problem for the two parameters you obtain the following first order conditions, which are the same as we obtained before multiplied by $-2n$, and, therefore, are solved for the same values,

$$-2\sum_{i=1}^{n} \; y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \; = 0$$

$$-2\sum_{i=1}^{n} x_i \; y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \; = 0$$

OLS: Alternate approach

- The above equations are called normal equations.
- See Appendix 2A for a formal proof that the solution of this system of two equations in two unknowns gives us a minimum of the objective function.

OLS: Alternate approach

- Why not to minimize some other function of the residuals, such as the absolute values of the residuals?.
- Mainly two reasons:

1. Some functions are meaningless, e.g. sum of residuals.
2. Other functions, e.g. sum of the absolute values of residuals, are more difficult to handle.

- OLS is simple and has good properties.

Algebraic Properties of OLS

- By definition, each fitted value $\hat{y}_i$ is on the OLS regression line. The OLS residual associated with observation i, $\hat{u}_i$, is the difference between y_i and its fitted value.
- If $\hat{u}_i$ is positive, the regression line underpredicts y_i; if $\hat{u}_i$ is negative, the regression line overpredicts y_i.
- The ideal case for observation i is $\hat{u}_i = 0$, but usually none of the data points will lie on the OLS regression line.

Algebraic Properties of OLS

- The sum of the OLS residuals is zero.
- Thus, the sample average of the OLS residuals is zero as well.
- The sample covariance between the regressor and the OLS residuals is zero.
- The OLS regression line always goes through the mean of the sample.

Algebraic Properties (precise)

(1) $\Sigma_{i=1}^{n}\hat{u}_i = 0 \qquad \Rightarrow \qquad \bar{\hat{u}} = \dfrac{\Sigma_{i=1}^{n}\hat{u}_i}{n} = 0$

(2) $\Sigma_{i=1}^{n}x_i\hat{u}_i = 0$

(3) $(\bar{y}, \bar{x})$ is on the regression line $\Rightarrow \quad \bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}$

Algebraic Properties (precise)

Writing $\quad y_i = \hat{y}_i + \hat{u}_i \quad$ we have

$$(1) \qquad \Sigma_{i=1}^{n} \hat{u}_i = 0 \qquad \Rightarrow \qquad \bar{y} = \bar{\hat{y}}$$

$$(1)+(2) \qquad \left.\begin{array}{l} \Sigma_{i=1}^{n} \hat{u}_i = 0 \\ \Sigma_{i=1}^{n} x_i \hat{u}_i = 0 \end{array}\right\} \qquad \Rightarrow \qquad \Sigma_{i=1}^{n} \hat{y}_i \hat{u}_i = 0$$

this last one implies that the sample covariance between fitted values, $\hat{y}_i$, and residuals, $\hat{u}_i$, is zero.

Algebraic Properties

- Thinking of each observation as being made up of an explained part, and an unexplained part, $y_i = \hat{y}_i + \hat{u}_i$, we can view OLS as decomposing each y_i into two parts, a fitted value and a residual. The fitted values and residuals are uncorrelated in the sample.

Sum of Squares Decomposition

- Define:

1. Total Sum of Squares (SST)

$$\mathrm{SST} \equiv \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2$$

2. Explained Sum of Squares (SSE)

$$\mathrm{SSE} \equiv \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y} \right)^2$$

3. Residual Sum of Squares (SSR)

$$\mathrm{SSR} \equiv \sum_{i=1}^{n} \hat{u}_i^2$$

Sum of Squares Decomposition

- SST is a measure of the total sample variation in the y_i.
- It can be shown that total variation in y, SST, can always be expressed as the sum of the explained variation, SSE, and the unexplained variation, SSR. Thus

$$\text{SST} = \text{SSE} + \text{SSR}$$

Proof that SST = SSE + SSR

$$
\begin{aligned}
\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2 &= \sum_{i=1}^{n}\left[(y_i-\hat{y}_i)+(\hat{y}_i-\bar{y})\right]^2 \\
&= \sum_{i=1}^{n}\left[\hat{u}_i+(\hat{y}_i-\bar{y})\right]^2 \\
&= \sum_{i=1}^{n}\hat{u}_i^2+2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\bar{y})+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \\
&= \mathrm{SSR}+\underbrace{\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\bar{y})}_{=0}+\mathrm{SSE}
\end{aligned}
$$

Given the above properties, so SST = SSE + SSR

Goodness-of-Fit

- How do we think about how well our sample regression line fits our sample data?
- We can compute the fraction of the total sum of squares (SST) that is explained by the model (SSE), call this the R-squared, R^2, of regression:

$$R^2 = SSE/SST = 1 - SSR/SST$$

Goodness-of-Fit

- $100.R^2$ is the percentage of the sample variation in *y* that is explained by *x* (the model).
- $R^2 \in [0,1]$
- If $R^2 = 1$, then we have a perfect fit, $\hat{u}_i = 0$ for all observations.
- If $R^2 = 0$, or close to zero, then we have a poor fit: very little variation in *y* is explained by *x*.

Goodness-of-Fit

- It can be shown that R^2 is equal to:

1. The square of the sample correlation coefficient between y_i and $\hat{y}_i$.
2. The square of the sample correlation coefficient between y_i and x_i.

Please show this as an exercise!.

Goodness-of-Fit

- Because we want high explanatory power for our models, we look, other things equal, for high R^2 in our regressions.
- However, in the social sciences, low R^2 in regression equations are not uncommon, especially for cross-section analysis.
- It is worth emphasizing now that a seemingly low R^2 does not necessarily mean that an OLS regression equation is useless.
- Goodness of fit is not the only feature we look for in a regression equation.
- More on this in the multiple regression analysis.

Using software for OLS regressions

- Now that we have derived the formula for calculating the OLS estimates of our parameters given a sample, you will be happy to know you don't have to compute them by hand.

- Regression packages, like Eviews, TSP, RATS, Stata …, will do the job for you.

Units of Measurement and Functional Form

- Two important issues in applied economics are:

1. Understanding how changing the units of measurement of the dependent and/or independent variables affects OLS estimates, and
2. Knowing how to incorporate popular functional forms used in economics into regression analysis.

Units of Measurement

- If x is multiplied/divided by a constant, $c \neq 0$, then the OLS slope is divided/multiplied by the same constant, c.
- Changing the units of measurement of x only does not affect the intercept.
- If y is multiplied/divided by a constant, $c \neq 0$, then the OLS slope and intercept are both multiplied/divided by the same constant, c.

Changing the Origin

- If add a constant *d* to *x* and/or *y* then the OLS slope is not affected.
- However changing the origin of either *x* and/or *y* affects the intercept of the regression.
- Eventually note that the goodness of fit, R^2, is invariant to changes in the units of *x* and/or *y*, and also to the origin of the variables.
- Try exercise 2.9 to practice this!

Functional Form

- Linear relationships are not general enough for all economic applications. However, we can incorporate many nonlinearities (in variables) into simple regression analysis by appropriately redefining the dependent and independent variables.

  Example, $y = \beta_0 + \beta_1 x^2 + u$

- We consider some possibilities that often appear in applied work.

Proportions and Percentages

- Remember that:

1. Proportional change: $$\frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$$
2. Percentage change: $$100.\frac{\Delta x}{x_0} = \%\Delta x$$
3. Elasticity: $$\frac{\Delta y}{\Delta x}.\frac{x_0}{y_0} = \frac{\%\Delta y}{\%\Delta x}$$

Proportions and Percentages

4. Changes in logarithms:

$$\Delta \log(x) = \log(x_1) - \log(x_0) \approx \frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x_0}$$

Hence,

$$100.\Delta \log(x) \approx \%.\Delta x$$

A Linear Model for log(y)

- Consider the model

$$\log(y) = \beta_0 + \beta_1 x + u$$

- What is the meaning of β_1 in this model?
- If $\Delta u = 0$, then x has a linear effect on log(y):

$$\Delta \log(y) = \beta_1 \Delta x$$

  or,

$$\%\Delta y = (100.\beta_1).\Delta x$$

  i.e. $100.\beta_1$ is the percentage change in y by unit of x.

A Constant Elasticity Model

- Consider the model

$$\log(y) = \beta_0 + \beta_1 \log(x) + u$$

- What is the meaning of β_1 in this model?
- If $\Delta u = 0$, then $\log(x)$ has a linear effect on $\log(y)$:

$$\Delta\log(y) = \beta_1 \Delta\log(x) \quad \Leftrightarrow \quad \%\Delta y = \beta_1 . \%\Delta x$$

i.e. β_1 is the elasticity of y with respect to x.

Functional Forms Involving logs

Model	Dependent Variable	Independent Variable	Interpretation of β_1
level-level	y	x	$\Delta y = \beta_1 . \Delta x$
level-log	y	$\log(x)$	$\Delta y = (\beta_1/100).\%\Delta x$
log-level	$\log(y)$	x	$\%\Delta y = (100.\beta_1).\Delta x$
log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1.\%\Delta x$

Functional Form

- Important: While the mechanics of simple regression does not depend on how *y* and *x* are defined, the interpretation of the coefficients does depend on their definitions.

The Meaning of "Linear" Regression

- We have claimed that our regression model, $y = \beta_0 + \beta_1 x + u$, is linear. This is true in two senses:

1. It is linear in the variables involved, y and x. And we have just seen how this can be relaxed to some extend.
2. It is linear in a more fundamental sense. The key of "linearity" is that it is linear in the parameters, β_0 and β_1, and u is additive.

The Meaning of "Linear" Regression

- For OLS to be applied as an estimation method to a given model, this should be linear in parameters.
- Plenty of models cannot be cast as linear regression models because they are not linear in parameters.

  i.e.

$$y = \frac{1}{\beta_0 + \beta_1 x} + u \qquad \text{or} \qquad y = \beta_0 e^{\beta_1 x} + u$$

- Estimation of such models takes us into the topic of nonlinear regression.

Statistical Properties of OLS

- We defined de population model
  $y = \beta_0 + \beta_1 x + u$, and we claimed that the key assumption for the simple regression analysis to be useful is that $E(u|x) = 0$.
- We now return to the population model and study the statistical properties of OLS estimators, $\hat{\beta}_0$ and $\hat{\beta}_1$, considered as estimators of the population parameters, β_0 and β_1.

Assumptions

- **SLR.1: LINEAR IN PARAMETERS**

  The population model is linear in parameters and given by

$$y = \beta_0 + \beta_1 x + u$$

Assumptions

- **SLR.2: RANDOM SAMPLING**

  We have a random sample from of size n, $\{(y_i,x_i): i = 1,2,3,\ldots,n\}$, from the population model.

  Thus we can write the population model in terms of the sample,

$$y_i = \beta_0 + \beta_1 x_i + u_i\,, \quad i = 1,2,3,\ldots,n$$

Assumptions

- **SLR.3: ZERO CONDITIONAL MEAN**

$$\mathrm{E}(u|x) = 0$$

For a random sample, this assumption implies that

$$\mathrm{E}(u_i|x_i) = 0\,, \quad i = 1,2,3,\ldots,n$$

NOTE: Derivations will be conditional on the sample values, x's.

Assumptions

- **SLR.4:** **SAMPLE VARIATION IN THE INDEPENDENT VARIABLE**

  The independent variables x_i, $i = 1,2,3,\ldots,n$, are not all equal to the same constant.

$$\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2 > 0$$

  This requires some variation in the population.

Unbiasedness of OLS

- We focus on the slope parameter, β_1.
- In order to think about unbiasedness, we need to rewrite our estimator in terms of the population parameter.
- Start with a simple rewrite of the OLS formula as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i - \bar{x} \quad y_i - \bar{y}}{\sum_{i=1}^{n} x_i - \bar{x} \;^{2}} = \frac{\sum_{i=1}^{n} x_i - \bar{x} \;\; y_i}{\mathrm{SST}_x}$$

where $\mathrm{SST}_x = \sum_{i=1}^{n} x_i - \bar{x} \;^{2}$

Unbiasedness of OLS

- Substituting $y_i = \beta_0 + \beta_1 x + u$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(\beta_0 + \beta_1 x_i + u_i\right)}{\mathrm{SST}_x}$$

$$= \beta_1 + \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right) u_i}{\mathrm{SST}_x}$$

The estimator equals the population slope, β_1, plus a term that is a linear combination in the errors, $\{u_1,\ldots,u_n\}$.

Unbiasedness of OLS

- Remember that

$$\sum_{i=1}^{n} \; x_i - \bar{x} \;\; = 0$$

$$\sum_{i=1}^{n} \; x_i - \bar{x} \;\; x_i = \sum_{i=1}^{n} \; x_i - \bar{x} \;\;\; x_i - \bar{x} \;\; = \sum_{i=1}^{n} \; x_i - \bar{x} \;\; {}^{2}$$

let $\; d_i = \; x_i - \bar{x} \;$, so that $\; \hat{\beta}_1 = \beta_1 + \left(\dfrac{1}{\mathrm{SST}_x} \right) \sum_{i=1}^{n} d_i u_i$

- Taking expectations, conditionally on the *x*'s, we can prove the first important statistical property of OLS.

Unbiasedness of OLS

- **THEOREM 2.1 UNBIASEDNESS OF OLS**

  Under assumptions SLR.1 to SLR.4

$$E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad E(\hat{\beta}_1) = \beta_1$$

  PROOF:

$$E(\hat{\beta}_1) = \beta_1 + E\left[\left(\frac{1}{\mathrm{SST}_x}\right)\sum_{i=1}^{n} d_i u_i\right] = \beta_1 + \left(\frac{1}{\mathrm{SST}_x}\right)\sum_{i=1}^{n} E\ \ d_i u_i$$

$$= \beta_1 + \left(\frac{1}{\mathrm{SST}_x}\right)\sum_{i=1}^{n} d_i E(u_i) = \beta_1 + \left(\frac{1}{\mathrm{SST}_x}\right)\sum_{i=1}^{n} d_i . 0 = \beta_1$$

Unbiasedness Summary

- The OLS estimates of β_0 and β_1 are unbiased.
- The proof of unbiasedness depends on our 4 assumptions – if any assumption fails, then OLS is not necessarily unbiased.
- Remember unbiasedness is a description of the estimator – in a given sample we may be "near" or "far" from the true parameter, but its distribution will be centered at the population parameter.

Variance of the OLS Estimators

- Now we know that the sampling distribution of our estimator is centered around the true parameter.
- How spread out this distribution is? This will be a measure of uncertainty.
- It is much easier to think about this variance under an additional assumption.

Assumptions

- **SLR.5: HOMOSKEDASTICITY**

$$\mathrm{Var}(u|x) = \sigma^2$$

Variance of the OLS Estimators

- The homoskedasticity assumption is quite distinct from the zero conditional mean assumption, $E(u|x) = 0$. SLR.3 involves the expected value of u, while SLR.5 concerns the variance of u.
- Homoskedasticity plays no role in showing that $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased.
- We add SLR.5 because it simplifies the variance calculations and because it implies that OLS has certain efficiency properties.

Variance of the OLS Estimators

- $\text{Var}(u|x) = \sigma^2 = \text{E}(u^2|x) - [\text{E}(u|x)]^2$
- $\text{E}(u|x) = 0$, so $\sigma^2 = \text{E}(u^2|x) = \text{E}(u^2) = \text{Var}(u)$
- Thus σ^2 is also the unconditional variance, called the error variance.
- σ, the square root of the error variance, is called the standard deviation of the error.

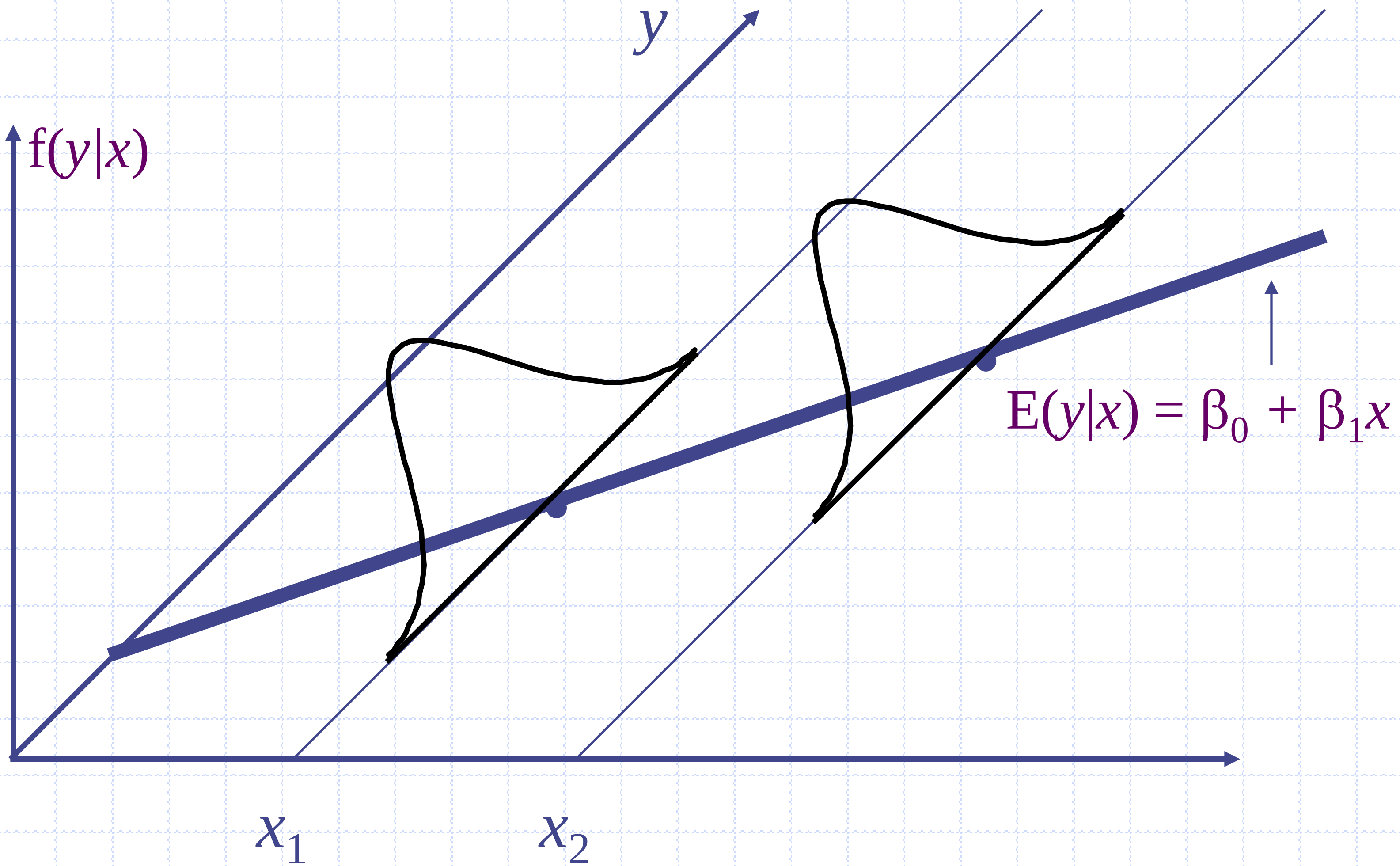
Variance of the OLS Estimators

- We can say:

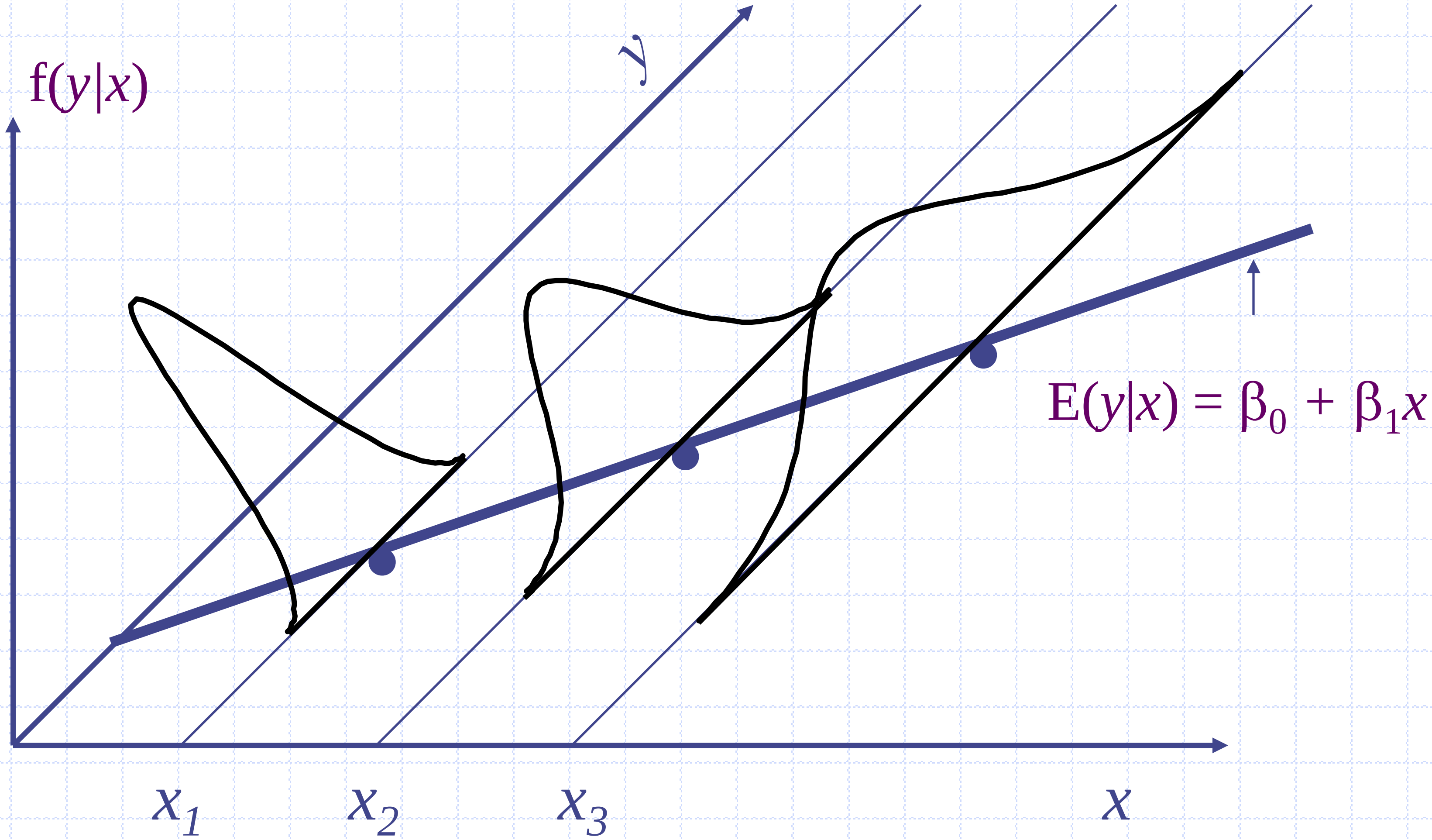
$$E(y|x) = \beta_0 + \beta_1 x \quad \text{and} \quad Var(y|x) = \sigma^2.$$

- So, the conditional expectation of y given x is linear in x, but the variance of y given x is constant.
- When $Var(u|x)$ depends on x, the error term is said to exhibit heteroskedasticity. Since $Var(u|x) = Var(y|x)$, heteroskedasticity is present whenever $Var(y|x)$ is a function of x.

Homoskedastic Case



Heteroskedastic Case



Variance of OLS estimators

- **THEOREM 2.2 SAMPLING VARIANCES OF OLS ESTIMATORS**

  Under assumptions SLR.1 to SLR.5

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \qquad \text{and} \qquad Var(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

  where these are conditional on the sample values $\{x_1, \ldots, x_n\}$

Variance of OLS estimators

- **PROOF:**

$$Var(\hat{\beta}_1) = E\left[\ \hat{\beta}_1 - E(\hat{\beta}_1)\ ^2\right] = E\left[(\hat{\beta}_1 - \beta_1)^2\right]$$

$$= E\left[\left(\left(\frac{1}{\mathrm{SST}_x}\right)\sum_{i=1}^{n} d_i u_i\right)^2\right] = \left(\frac{1}{\mathrm{SST}_x}\right)^2 E\left[\left(\sum_{i=1}^{n} d_i u_i\right)^2\right]$$

$$= \left(\frac{1}{\mathrm{SST}_x}\right)^2 \left(\sum_{i=1}^{n} d_i^2 E(u_i^2)\right) = \sigma^2 \left(\frac{1}{\mathrm{SST}_x}\right)^2 \underbrace{\left(\sum_{i=1}^{n} d_i^2\right)}_{\mathrm{SST}_x}$$

$$= \sigma^2 \left(\frac{1}{\mathrm{SST}_x}\right)^2 \mathrm{SST}_x = \frac{\sigma^2}{\mathrm{SST}_x}$$

Variance of OLS: Summary

- The larger the error variance, σ^2, the larger the variance of the slope estimate.
- The larger the variability in the x_i's, the smaller the variance of the slope estimate.
- As a result, a larger sample size should decrease the variance of the slope estimate.
- Problem: the error variance, σ^2, is unknown.

Estimating the Error Variance

- We don't know what the error variance, σ^2, is, and we cannot estimate it from the errors, u_i, because we don't observe the errors.
- $\sigma^2 = E(u^2)$, so an unbiased "estimator" would be $n^{-1}\Sigma_{i=1}^{n} u_i^2$.
- Unfortunately, this is not a true estimator, because we don't observe the errors u_i. But, we do have estimates of the u_i, namely the OLS residuals $\hat{u}_i$.

Estimating the Error Variance

- The relation between errors and residuals is given by

$$\hat{u}_i = y_i - \hat{y}_i = \beta_0 + \beta_1 x_i + u_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

$$= u_i - \left(\hat{\beta}_0 - \beta_0\right) - \left(\hat{\beta}_1 - \beta_1\right) x_i$$

- Hence $\hat{u}_i$ is not the same as u_i, although the difference between them does have an expected value of zero.

Estimating the Error Variance

- If we replace the errors with the OLS residuals, we have $n^{-1}\Sigma_{i=1}^{n}\hat{u}_i^2 = \text{SSR}/n$
- This is a true estimator, because it gives a computable rule for any sample of the data, *x* and *y*.
- However, this estimator is biased, essentially because it does not account for two restrictions that must be satisfied by the OLS residuals, $n^{-1}\Sigma_{i=1}^{n}\hat{u}_i = 0$ and $n^{-1}\Sigma_{i=1}^{n}x_i\hat{u}_i = 0$

Estimating the Error Variance

- One way to view these restrictions is this: If we know $n - 2$ of the residuals, we can get the other two residuals by using the restrictions implied by the moment conditions.
- Thus, there are only $n - 2$ degrees of freedom in the OLS residuals, as opposed to n degrees of freedom in the errors.

Estimating the Error Variance

- The unbiased estimator of σ^2 that we will use makes a degrees of freedom adjustment:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2} = \frac{\text{SSR}}{n-2}$$

- **THEOREM 2.3 UNBIASED ESTIMATOR OF σ^2**

  Under assumptions SLR.1 to SLR.5

$$E(\hat{\sigma}^2) = \sigma^2$$

Estimating the Error Variance

- If $\hat{\sigma}^2$ is plugged into the variance formulas we then have unbiased estimators of $Var(\hat{\beta}_1)$ and $Var(\hat{\beta}_0)$
- The natural estimator of σ is $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ and is called the **standard error of the regression**.
- Since $sd(\hat{\beta}_1) = \sigma / \sqrt{SST_x}$, its natural estimator is

$$se(\hat{\beta}_1) = \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Estimating the Error Variance

- Note that $se(\hat{\beta}_1)$, the **standard error** of $\hat{\beta}_1$, is view as a random variable when we think of running OLS over different samples; this is because $\hat{\sigma}$ varies with different samples.

- The standard error of any estimate gives us an idea of how precise the estimator is.

Regression Through the Origin

- If we force the regression line to pass through the point (0,0) we are constraining the intercept to be zero.
- This is called a regression through the origin.
- This is not done very often since, among other problems, when $\beta_0 \neq 0$, then the slope estimate will be biased.
- Summary: always include an intercept in your regressions.

Appendix: $R^2 = r^2_{y,\hat{y}}$

$$\left.\begin{aligned} R^2 &= \frac{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2} \\ r^2_{y,\hat{y}} &= \frac{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(\hat{y}_i-\bar{\hat{y}}\right)\right)^2}{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2\right)\left(\sum_{i=1}^{n}\left(\hat{y}_i-\bar{\hat{y}}\right)^2\right)} \end{aligned}\right\}$$

but since $\bar{y} = \bar{\hat{y}}$

$$r^2_{y,\hat{y}} = \frac{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(\hat{y}_i-\bar{y}\right)\right)^2}{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)^2\right)\left(\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2\right)}$$

Appendix: $\mathrm{R}^2 = r^2_{y,\hat{y}}$

$$
\begin{aligned}
\boxed{\sum_{i=1}^{n} \; y_i - \bar{y} \quad \hat{y}_i - \bar{y}} &= \sum_{i=1}^{n} y_i \;\; \hat{y}_i - \bar{y} \\
&= \sum_{i=1}^{n} \; \hat{y}_i + \hat{u}_i \quad \hat{y}_i - \bar{y} \\
&= \sum_{i=1}^{n} \hat{y}_i \;\; \hat{y}_i - \bar{y} \;\; + \sum_{i=1}^{n} \hat{u}_i \;\; \hat{y}_i - \bar{y} \\
&= \sum_{i=1}^{n} \hat{y}_i \;\; \hat{y}_i - \bar{y} \;\; + \underbrace{\sum_{i=1}^{n} \hat{u}_i \hat{y}_i}_{=0} - \bar{y} \underbrace{\sum_{i=1}^{n} \hat{u}_i}_{=0} \\
&= \sum_{i=1}^{n} \; \hat{y}_i - \bar{y} \quad \hat{y}_i - \bar{y} \;\; = \boxed{\sum_{i=1}^{n} \; \hat{y}_i - \bar{y} \;\; ^2}
\end{aligned}
$$

Appendix: $\mathrm{R}^2 = r^2_{y,\hat{y}}$

$$\boxed{\mathrm{R}^2} = \frac{\sum_{i=1}^{n} \hat{y}_i - \bar{y}\;^2}{\sum_{i=1}^{n} y_i - \bar{y}\;^2} =$$

$$= \frac{\sum_{i=1}^{n} \hat{y}_i - \bar{y}\;^2}{\sum_{i=1}^{n} y_i - \bar{y}\;^2} \cdot \frac{\sum_{i=1}^{n} \hat{y}_i - \bar{y}\;^2}{\sum_{i=1}^{n} \hat{y}_i - \bar{y}\;^2} = \frac{\left(\sum_{i=1}^{n} y_i - \bar{y} \quad \hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} y_i - \bar{y}\;^2 \sum_{i=1}^{n} \hat{y}_i - \bar{y}\;^2} = \boxed{r^2_{y,\hat{y}}}$$

Appendix: $\mathrm{R}^2 = r_{y,x}^2$

$$\left.\begin{array}{l} \mathrm{R}^2 = \dfrac{\sum\limits_{i=1}^{n} \hat{y}_i - \bar{y} \quad ^2}{\sum\limits_{i=1}^{n} y_i - \bar{y} \quad ^2} \\[2em] r_{y,x}^2 = \dfrac{\left(\sum\limits_{i=1}^{n} y_i - \bar{y} \quad x_i - \bar{x} \right)^2}{\left(\sum\limits_{i=1}^{n} y_i - \bar{y} \quad ^2 \right)\left(\sum\limits_{i=1}^{n} x_i - \bar{x} \quad ^2 \right)} \end{array}\right\} \text{but} \quad \left\{ \begin{array}{l} \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \\ \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \end{array} \right\} \Rightarrow \hat{y}_i - \bar{y} = \hat{\beta}_1 \quad x_i - \bar{x}$$

Appendix: $\mathrm{R}^2 = r_{y,x}^2$

$$\boxed{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2}=\hat{\beta}_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

$$=\underbrace{\frac{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(x_i-\bar{x}\right)\right)^2}{\left(\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right)^2}}_{\hat{\beta}_1^2}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$$

$$=\boxed{\frac{\left(\sum_{i=1}^{n}\left(y_i-\bar{y}\right)\left(x_i-\bar{x}\right)\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}$$

Appendix: $\mathrm{R}^2 = r_{y,x}^2$

$$\boxed{\mathrm{R}^2} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} =$$

$$= \frac{\left(\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) \right)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (x_i - \bar{x})^2} = \boxed{r_{y,x}^2}$$