Multiple Regression Analysis

Caio Vigo

The University of Kansas
Department of Economics

Summer 2019

These slides were based on *Introductory Econometrics* by Jeffrey M. Wooldridge (2015)

Topics

Motivation:

• With a simple linear regression model we learned a model in which a **single** independent variable x explains (or affect) a dependent variable y.

• If we add more factors to our model that are useful for explaining y, then more of the variation in y can be explained.

We can build better models for predicting the dependent variable.

Motivation for Multiple Regression

- Recall the *log(wage)* example.

Example: *log(wage)*

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

- Might be the case that there are factors in u affecting y.
- For instance intelligence could help to explain *wage*.

- Let's use a **proxy** for it: IQ.
- By explicitly including IQ in the equation, we can take it out of the error term.
- Consider the following extension of the *log(wage)* example:

Example: *log(wage)* (extension)

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + u$$

The Model with 2 Independent Variable

Generally, we can write a model with two independent variables as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

where

β_0 is the intercept,
β_1 measures the change in y with respect to x_1, *holding other factors fixed*,
β_2 measures the change in y with respect to x_2, *holding other factors fixed*

The Model with 2 Independent Variable

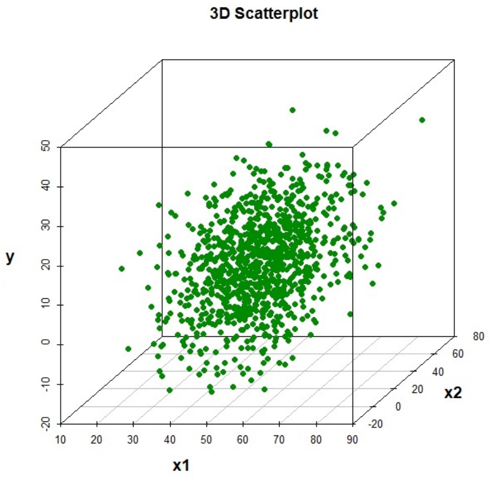
- In the model with two explanatory variables, the key assumption about how u is related to x_1 and x_2 is:

$$E(u|x_1, x_2) = 0.$$

- For any values of x_1 and x_2 in the population, the average unobservable is equal to zero.

- The value zero is not important because we have an intercept, β_0 in the equation.

The Model with 2 Independent Variable



The Model with 2 Independent Variable

- In the wage equation, the assumption is $E(u|educ, IQ) = 0$.

- Now u no longer contains intelligence, and so this condition has a better chance of being true.

- Recall that in the simple regression, we had to assume IQ and $educ$ are unrelated to justify leaving IQ in the error term.

- Other factors, such as workforce experience and "motivation," are part of u. Motivation is very difficult to measure. Experience is easier:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 \mathit{exper} + u.$$

The Model with k Independent Variables

- The **multiple linear regression model** can be written in the population as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

where,

β_0 is the **intercept**,
β_1 is the parameter associated with x_1,
β_2 is the parameter associated with x_2, and so on.

- Contains $k + 1$ **(unknown) population parameters.**

- We call $\beta_1, \ldots, \beta_k$ the **slope parameters**.

The Model with k Independent Variables

- Now we have **multiple explanatory** or **independent variables** $x's$.
- We still have **one explained** or **dependent variable** y.
- We still have **an error term**, u.

The Model with k Independent Variables

- **Advantage of multiple regression:** it can incorporate fairly general functional form relationships.

- Let $lwage = \log(wage)$:

$$lwage = \beta_0 + \beta_1 educ + \beta_2 IQ + \beta_3 \textit{exper} + \beta_4 \textit{exper}^2 + u,$$

so that *exper* is allowed to have a quadratic effect on $lwage$.

- Thus, $x_1 = educ$, $x_2 = IQ$, $x_3 = \textit{exper}$, and $x_4 = \textit{exper}^2$. Note that x_4 is a a *nonlinear* function of x_3.

The Model with k Independent Variables

- The key assumption for the general multiple regression model is:

$$E(u|x_1, ..., x_k) = 0$$

- We can make this condition closer to being true by "controlling for" more variables.

Topics

Mechanics and Interpretation of OLS

- Suppose we have x_1 and x_2 $(k = 2)$ along with y.
- We want to fit an equation of the form:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

given data $\{(x_{i1}, x_{i2}, y_i) : i = 1, ..., n\}$.

- Sample size $= n$.

Labels and indexing

Now the explanatory variables have two subscripts:

- i = observation number
- j = labels for particular variables (it is the second subscript - 1 and 2 in this case)

For example:

$$\begin{aligned} x_{i1} &= educ_i \quad , i = 1, 2, \ldots, n \\ x_{i2} &= IQ_i \quad , i = 1, 2, \ldots, n \end{aligned}$$

Least Squares Method

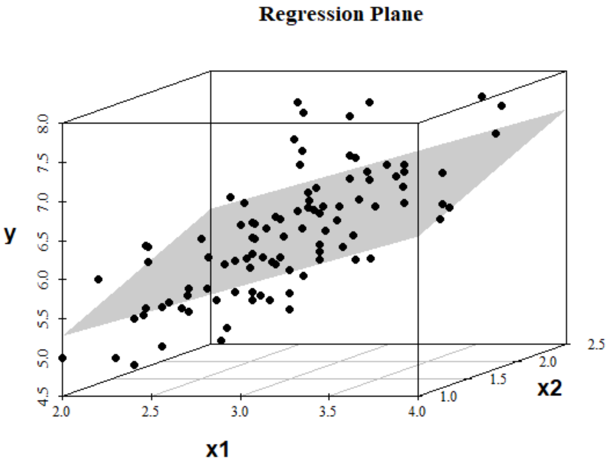
- As in the simple regression case, different ways to motivate OLS. We choose $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ (so three unknowns) to minimize the sum of squared residuals,

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$$

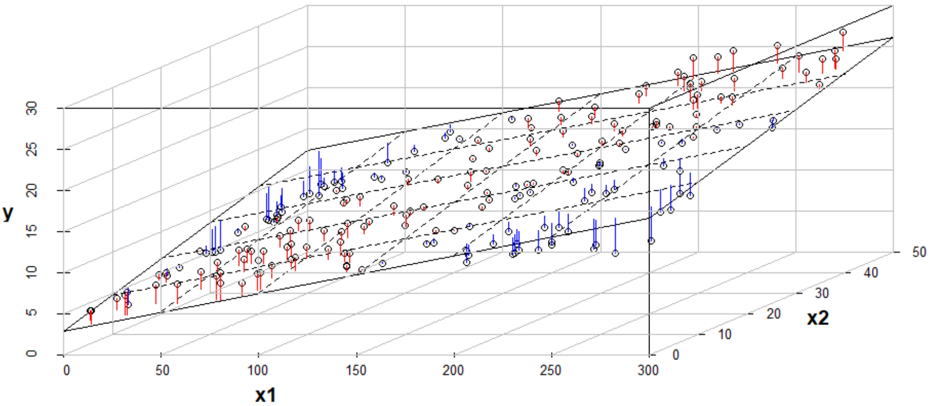
- The case with k independent variables is easy to state: choose the $k+1$ values $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_k$ to minimize

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})^2$$

Derivation of the OLS Estimator - Least Squares Method



Derivation of the OLS Estimator - Least Squares Method



Derivation of the OLS Estimator - Least Squares Method

- The **OLS first order conditions** (solved with multivariable calculus) are the $k+1$ linear equations in the $k+1$ unknowns $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2, \ldots,$ $\hat{\beta}_k$:

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\vdots &= \vdots \\
\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) &= 0
\end{aligned}
$$

Derivation of the OLS Estimator - Least Squares Method

- As long as we add an assumption (**MLR.3** - we will see in the next topic),we can guarantee this system to have an unique solution.
- We will not find a closed solution to each β_j , for $j = 0, 1, 2, \ldots, k$.
- We can use matrix algebra to easily find the solution.

The OLS regression line is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k$$

Interpreting the OLS Regression Line

- The slope coefficients now explicitly have ceteris paribus interpretations.
- Consider $k = 2$:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

Then

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2$$

allows us to compute how predicted y changes when x_1 and x_2 change by any amount.

Interpreting the OLS Regression Line

- What if we “hold x_2 fixed,” that is, its change is zero, $\Delta x_2 = 0$?

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 \text{ if } \Delta x_2 = 0$$

In particular,

$$\hat{\beta}_1 = \frac{\Delta \hat{y}}{\Delta x_1} \text{ if } \Delta x_2 = 0$$

In other words, $\hat{\beta}_1$ is the slope of $\hat{y}$ with respect to x_1 when x_2 is held fixed.

Interpreting the OLS Regression Line

- Similarly,

$$\Delta \hat{y} = \hat{\beta}_2 \Delta x_2 \text{ if } \Delta x_1 = 0$$

and

$$\hat{\beta}_2 = \frac{\Delta \hat{y}}{\Delta x_2} \text{ if } \Delta x_1 = 0$$

- We call $\hat{\beta}_1$ and $\hat{\beta}_2$ **partial effects** or **ceteris paribus effects**.

Terminology

We say that $\hat{\beta}_0$, $\hat{\beta}_1, ...,$ $\hat{\beta}_k$ are the OLS estimates from the regression

$$y \text{ on } x_1, x_2, ..., x_k$$

or

$$y_i \text{ on } x_{i1}, x_{i2}, ..., x_{ik}, \ \ i = 1, ..., n$$

when we want to emphasize the sample being used.

Interpreting the OLS Regression Line

- Recall the **wage** example:

Example (Wage)

$$\begin{aligned} \widehat{wage} &= -0.90 + 0.54\ educ \\ n &= 526, \qquad R^2 = .16 \end{aligned}$$

- Then we did:

$$\log(wage) = \beta_0 + \beta_1 educ + u$$

Interpreting the OLS Regression Line

$$\begin{aligned} \widehat{lwage} &= 0.58 + .08\ educ \\ n &= 526,\ R^2 = .19 \end{aligned}$$

- Let's write a multiple regression model:

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

Interpreting the OLS Regression Line - R Output

	Dependent variable:
	lwage
educ	0.092^{***}
	(0.007)
exper	0.004^{**}
	(0.002)
tenure	0.022^{***}
	(0.003)
Constant	0.284^{***}
	(0.104)
Observations	526
R^2	0.316
Adjusted R^2	0.312
Residual Std. Error	0.441 (df = 522)
F Statistic	80.391^{***} (df = 3; 522)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Interpreting the OLS Regression Line

$$\begin{aligned} \widehat{lwage} &= .284 + .092\ educ + .004\ exper + .022\ tenure \\ n &= 526,\ R^2 = .32 \end{aligned}$$

Interpretation:

• $.092$ means that, holding *exper* and *tenure* fixed, another year of education is predicted to increase *log(wage)* by $.092$, i.e., 9.2% **increase in *wage*.**

• Alternatively, we can take two people, A and B, with the *same exper* and *tenure*. Suppose person B has one more year of schooling than person A. Then we predict B to have a wage that is 9.2% higher.

What Does it Mean to "Hold Other Factors Fixed"?

• The power of multiple regression analysis is that it provides the *ceteris paribus* interpretation, even though the data have **not** been collected in a *ceteris paribus* fashion.

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_1 tenure + u$$

• Using the multiple regression model for wage as an example, it may seem that we actually went out and sampled people with the same *exper* and *tenure*.

• It's not the case. It's a random sample.

Fitted Values and Residuals

- For each i,

$$\begin{aligned}\hat{y}_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik} \\ \hat{u}_i &= y_i - \hat{y}_i\end{aligned}$$

OLS Fitted Values and Residuals

(1) The sample average of the residuals is zero, i.e., $\sum_{i=1}^{n} \hat{u}_i = 0$. This implies $\bar{y} = \bar{\hat{y}}$.

(2) Each explanatory variable is uncorrelated with the residuals in the sample. This follows from the first order conditions. It implies that $\hat{y}_i$ and $\hat{u}_i$ are also uncorrelated.

(3) The sample averages always fall on the OLS regression line:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + ... + \hat{\beta}_k \bar{x}_k$$

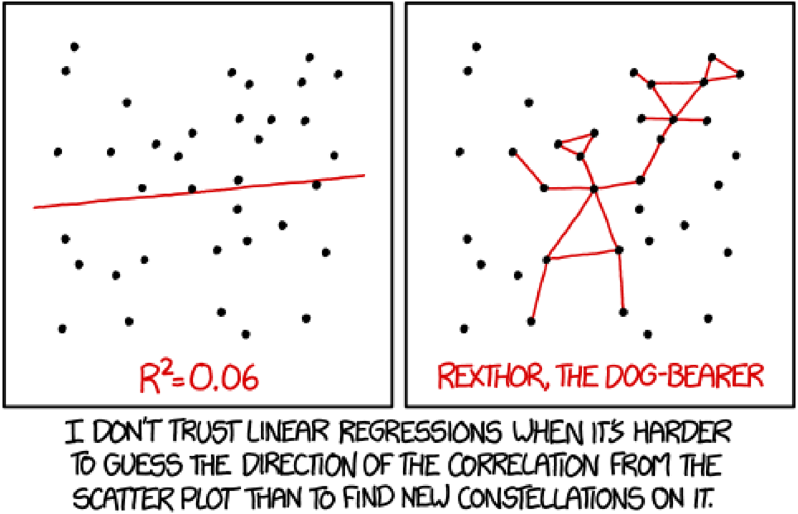
That is, if we plug in the sample average for each explanatory variable, the predicted value is the sample average of the y_i.

Goodness-of-Fit

$$R^2$$

... again

Goodness-of-Fit



Goodness-of-Fit

- As with simple regression, it can be shown that

$$SST = SSE + SSR$$

where SST, SSE, and SSR are the total, explained, and residual sum of squares.

- We define the R-squared as before:

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

Goodness-of-Fit

- Recall, $0 \leq R^2 \leq 1$

- Using the same set of data and the same dependent variable, the R^2 **can never fall when another independent variable is added to the regression.** And, it almost always goes up, at least a little.

- This means that, if we focus on R^2, we might include silly variables among the x_j.

- Adding another x cannot make SSR increase. The SSR falls unless the coefficient on the new variable is identically zero.

Topics

Assumptions

Assumption MLR.1 (Linear in Parameters)

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

where the β_j are the population parameters and u is the unobserved error.

Assumptions

Assumption MLR.2 (Random Sampling)

We have a random sample of size n from the population, $\{(x_{i1}, x_{i2}, ..., x_{ik}, y_i) : i = 1, ..., n\}$

- The data should be a representative sample from the population.

The 4 Assumptions for Unbiasedness

Assumption MLR.3 (No Perfect Collinearity)

In the sample (and, therefore, in the population), none of the explanatory variables is constant, and there are **no exact linear** relationships among them.

The 4 Assumptions for Unbiasedness

If an independent variable in a *Multiple Regression model* is an **exact linear combination** of the other independent variables, we say the model suffers from **perfect collinearity**, and it cannot be estimated by OLS.

- Under perfect collinearity, there are no unique OLS estimators. **R**, **Stata** and other regression packages will indicate a problem.

The 4 Assumptions for Unbiasedness

• We must rule out the (extreme) case that one (or more) of the explanatory variables is an exact *linear* function of the others.

Usually perfect collinearity arises from a **bad specification** of the population model.

• Assumption MLR.3 can only hold if $n \geq k+1$, that is, we must have at least as many observations as we have parameters to estimate.

The 4 Assumptions for Unbiasedness

- Suppose that $k = 2$ and $x_1 = educ$, $x_2 = exper$. If we draw our sample so that

$$educ_i = 2exper_i$$

for *every* i, then Assumption MLR.3 is violated.

- This is very unlikely unless the sample is small.

- In any realistic population there are plenty of people whose education level is not twice their years of workforce experience.

The 4 Assumptions for Unbiasedness

Do not include the same variable in an equation that is measured in different units.

Example: CEO Salary

In a CEO salary equation, it would make no sense to include firm sales measured in dollars along with sales measured in millions of dollars. There is no new information once we include one of these.

The 4 Assumptions for Unbiasedness

Be careful with functional forms! Suppose we start with a constant elasticity model of family consumption:

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + u$$

• How might we allow the elasticity to be nonconstant, but include the above as a special case? The following does *not* work:

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2 \log(inc^2) + u$$

because $\log(inc^2) = 2\log(inc)$, that is, $x_2 = 2x_1$, where $x_1 = \log(inc)$.

The 4 Assumptions for Unbiasedness

- Instead, we probably mean something like

$$\log(cons) = \beta_0 + \beta_1 \log(inc) + \beta_2[\log(inc)]^2 + u$$

which means $x_2 = x_1^2$. With this choice, x_2 is an exact *nonlinear* function of x_1, but this (fortunately) is allowed in MLR.3.

- Tracking down perfect collinearity can be harder when it involves more than two variables.

The 4 Assumptions for Unbiasedness

Example: Vote

$$voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend + u$$

where $expendA$ is campaign spending by candidate A, $expendB$ is spending by candidate B, and $totexpend$ is total spending. All are in thousands of dollars. Mechanically, the problem is that, by definition,

$$expendA + expendB = totexpend$$

which, of course, will also be true for any sample we collect.

The 4 Assumptions for Unbiasedness

- One of the three variables has to be dropped.

- The model makes no sense from a ceteris paribus perspective. For example, β_1 is suppose to measure the effect of changing $expendA$ on $voteA$, holding fixed $expendB$ and $totexpend$. But if $expendB$ and $totexpend$ are held fixed, $expendA$ cannot change!

- We would probably drop $totexpend$ and just use the two separate spending variables.

The 4 Assumptions for Unbiasedness

Key Point

Assumption MLR.3 does *not* say the explanatory variables have to be uncorrelated in the population or sample.

Nor does it say they cannot be "highly" correlated.

MLR.3 rules out *perfect correlation* in the sample, that is, correlations of ± 1.

• Multiple regression would be useless if we had to insist x_1, ..., x_k were uncorrelated in the sample (or population)!

• If the x_j were all pairwise uncorrelated, we could just use a bunch of simple regressions.

The 4 Assumptions for Unbiasedness

MLR.1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$
MLR.2: random sampling from the population
MLR.3: no perfect collinearity in the sample

- The last assumption ensures that the OLS estimators are unique and can be obtained from the first order conditions (minizing the sum of squared residuals).

- We need a final assumption for unbiasedness.

Assumption MLR.4 (Zero Conditional Mean)

$$E(u|x_1, x_2, ..., x_k) = 0 \text{ for all } (x_1, ..., x_k)$$

- Remember, the real assumption is $E(u|x_1, x_2, ..., x_k) = E(u)$: the average value of the error does not change across different slices of the population defined by $x_1, ..., x_k$.

- Setting $E(u) = 0$ essentially defines β_0.

The 4 Assumptions for Unbiasedness

If u is correlated with any of the x_j, **MLR.4** is violated.

- When Assumption MLR.4 holds, we say $x_1, ..., x_k$ are **exogenous explanatory variables**.

- If x_j is correlated with u, we often say x_j is an **endogenous explanatory variable**.

Theorem: Unbiasedness of OLS

Under Assumptions MLR.1 through MLR.4,

$$E(\hat{\beta}_j) = \beta_j,\ j = 0, 1, 2, ..., k$$

for any values of the population parameters β_j.. In other words, the OLS estimators are unbiased estimators of the population parameters.

KU

Topics

Homoskedasticity

- So far, we have assumed

MLR.1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
MLR.2: random sampling from the population
MLR.3: no perfect collinearity in the sample
MLR.4: $E(u|x_1, x_2, ..., x_k) = 0$

- Under MLR.3 we can compute the OLS estimates in our sample.

- The other assumptions then ensure that OLS is unbiased (conditional on the outcomes of the explanatory variables).

Homoskedasticity

- Now, our goal is to find $Var(\hat{\beta}_j)$.
- In order to do that we need to add another assumption: **homoskedasticity (constant variance).**
- Why should we add another assumption?
  1. Imposing this assumption, the OLS estimator has an important feature/property: **efficiency**.
  2. We can obtain simple formulas with it too.

Homoskedasticity

Assumption MLR.5 (Homoskedasticity)

The variance of the error, u, does not change with any of $x_1, x_2, ..., x_k$:

$$Var(u|x_1, x_2, ..., x_k) = Var(u) = \sigma^2$$

- What it is saying is that the variance of the unobservable, u, conditional on $x_1, x_2, ..., x_k$ is constant.

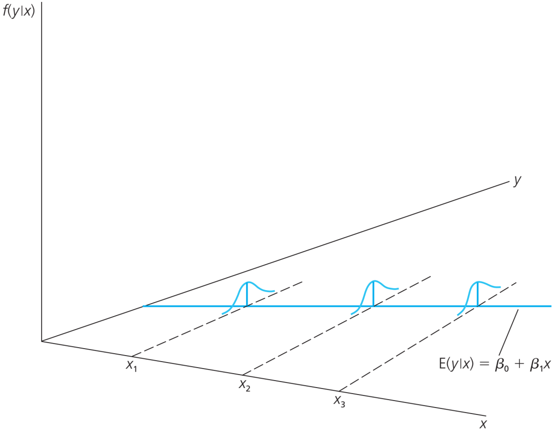
Homoskedasticity

- The homoskedasticity assumption is common in cross-section analysis. However there are many problems where it does not hold.
- For time series **hardly (!)** you can make this assumption.
- When $Var(u|x_1, x_2, ..., x_k)$ depends on x_j, the error term exhibits **heteroskedasticity** *(nonconstant variance)*
- Since $Var(u|x_1, x_2, ..., x_k) = Var(y|x_1, x_2, ..., x_k)$, we have **heteroskedasticity** when $Var(y|x_1, x_2, ..., x_k)$ is a function of x.

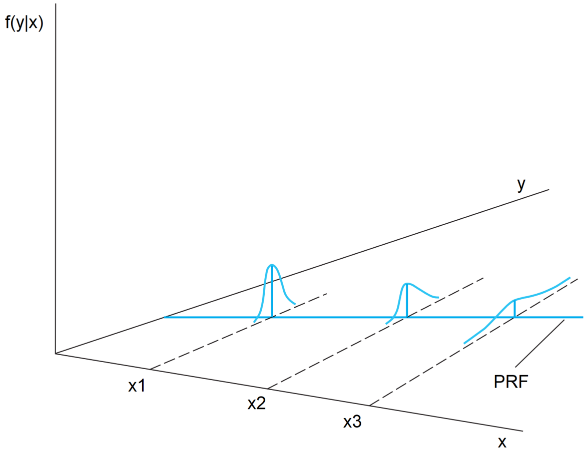
Homoskedasticity

- The homoskedasticity assumption plays no role in showing that $\hat{\beta}_j$ are unbiased.
- σ^2 is the unconditional variance of u.
- σ^2 : **error variance** or **disturbance variance.**
- $\sqrt{\sigma^2} = \sigma$: **standard deviation of the error**.

Homoskedasticity



Homoskedasticity



Homoskedasticity

- Assumptions **MLR.1** and **MLR.4** imply

$$E(y|x_1, x_2, ..., x_k) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

and when we add **MLR.5**,

$$Var(y|x_1, x_2, ..., x_k) = Var(u|x_1, x_2, ..., x_k) = \sigma^2$$

- Assumptions **MLR.1** through **MLR.5** are called the **Gauss Markov assumptions**.

Homoskedasticity

Gauss Markov assumptions

MLR.1: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$
MLR.2: random sampling from the population
MLR.3: no perfect collinearity in the sample
MLR.4: $E(u|x_1, x_2, ..., x_k) = 0$
MLR.5: $Var(u|x_1, x_2, ..., x_k) = Var(u) = \sigma^2$

The Variance of OLS

Recall, our goal is to find $Var(\hat{\beta}_j)$
(We will not find $Var(\hat{\beta}_0)$ - which has different formula)

- Let's define the total variation in x_j in the sample:

$$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

Notice that the R-squared can also be understood as the squared correlation between to variables.

- Let's define R-**squared** R_j^2:

a measure of correlation between x_j and the other explanatory variables (in the sample) is the R-squared from the regression:

$$x_{ij} \text{ on } x_{i1},\ x_{i2},\ \ldots, x_{i,j-1}, x_{i,j+1}, \ldots, x_{ik}$$

We are regressing x_j on all of the *other* explanatory variables.

Theorem: Sampling Variances of OLS Slope Estimators

Under Assumptions **MLR.1** to **MLR.5**, and condition on the values of the explanatory variables in the sample,

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \; j = 1, 2, ..., k.$$

• Clearly, all five Gauss-Markov assumptions are needed to ensure this formula is correct.

The Variance of OLS

- If,

$$Var(u|x_1, x_2, \ldots, x_k) = f(x_j)$$

- **Example:** On the white board.

- This violates **MLR.5**, and the standard variance formula is *generally* incorrect for **all** OLS estimators, not just $Var(\hat{\beta}_j)$.

- Is $R_j^2 = 1$ allowed? Answer: **No**.
- Any value $0 \leq R_j^2 < 1$ is permitted.
- **Multicollinearity** As R_j^2 gets closer to one, x_j is more linearly related to the other independent variables.

- The variance

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

has three components:

- σ^2
- SST_j
- $1 - R_j^2$

The Components of the OLS Variances

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

Factors Affecting $Var(\hat{\beta}_j)$:

(1) If the error variance $\sigma^2 \downarrow$,
$\Rightarrow Var(\hat{\beta}_j) \downarrow$ $\quad \Rightarrow Var(u|\mathbf{X}) \downarrow$ adding more explanatory variables

The Components of the OLS Variances

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1-R_j^2)}$$

Factors Affecting $Var(\hat{\beta}_j)$:

(2) If the $SST_j \uparrow$,
$Var(\hat{\beta}_j) \downarrow \Rightarrow$ the higher is the sample variation in x_j the better (increase the sample size n: SST_j is roughly a linear function of n).

The Components of the OLS Variances

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

Factors Affecting $Var(\hat{\beta}_j)$:

(3) As $R_j^2 \to 1$,
$Var(\hat{\beta}_j) \to \infty \Rightarrow R_j^2$ measures how linearly related x_j is to the other explantory variables.

KU

The Components of the OLS Variances

- We get the smallest variance for $\hat{\beta}_j$ when $R_j^2 = 0$:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j},$$

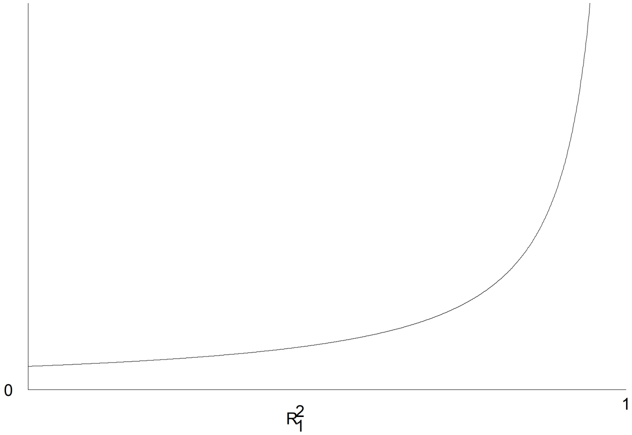
- If x_j is unrelated to all other independent variables $\Rightarrow$ easier to estimate its ceteris paribus effect on y.

- $R_j^2 \approx 0$ (uncommon).

- $R_j^2 \approx 1$ (more common) $\Rightarrow$ the estimate of β_j is not precise.

The Components of the OLS Variances

Figure: Graph of $Var(\hat{\beta}_1)$ as a function of R_1^2



The Components of the OLS Variances

Recall,

Multicollinearity: R_j^2 close to one. (problem of ...)
Perfect Collinearity: $R_j^2 = 1$ (not allowed under **MLR.1 - MLR.4**)

• Does multicollinearity (high correlation among two or more independent variables) violates any of the Gauss-Markov assumptions (including MLR.3.)?

Answer: **No.** Multicollinearity does not cause the OLS estimators to be biased. We still have $E(\hat{\beta}_j) = \beta_j$.

Estimating the Error Variance

Goal: We need to estimate σ^2.

- **Problem:** we don't observe u_i.
- We could use our residuals $\hat{u}_i$ (that we obtain when we run a regression) to find σ^2.

Estimating the Error Variance

- **Degrees of freedom:** With n observations and $k+1$ parameters, we only have

$$df = n - (k+1)$$

degrees of freedom. Recall we lose the $k+1$ df due to $k+1$ restrictions on the OLS residuals:

$$\begin{aligned} \sum_{i=1}^{n} \hat{u}_i &= 0 \\ \sum_{i=1}^{n} x_{ij}\hat{u}_i &= 0,\ j = 1, 2, ..., k \end{aligned}$$

Estimating the Error Variance

Estimator of σ^2

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{(n-k-1)} = \frac{SSR}{df}$$

- Regression packages (e.g. **R**) reports:
  - $\sqrt{\hat{\sigma}^2} = \hat{\sigma}$
  - **Names:** *Residual std. error, std. error of the regression, root mean squared error, standard error of the estimate, root mean squared error*

Note that SSR falls when a new explanatory variable is added, but df falls, too. So $\hat{\sigma}$ can increase or decrease when a new variable is added in multiple regression.

Estimating the Error Variance

Theorem: Unbiased Estimation of σ^2

Under the Gauss-Markov assumptions **MLR.1** through **MLR.5**

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

i.e., $\hat{\sigma}^2$ is an unbiased estimator of σ^2.

Standard Errors of the OLS Estimators

Goal: Now we want to find the **standard error** of each $\hat{\beta}_j$.

Standard deviation of $\hat{\beta}_j$

$$sd(\hat{\beta}_j) = \frac{\sigma}{\sqrt{SST_j(1-R_j^2)}}$$

Standard error of $\hat{\beta}_j$

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}$$

Standard Errors of the OLS Estimators

	Dependent variable:
	lwage
educ	0.092^{***}
	(0.007)
exper	0.004^{**}
	(0.002)
tenure	0.022^{***}
	(0.003)
Constant	0.284^{***}
	(0.104)
Observations	526
R^2	0.316
Adjusted R^2	0.312
Residual Std. Error	0.441 (df = 522)
F Statistic	80.391^{***} (df = 3; 522)
Note:	*p<0.1; **p<0.05; ***p<0.01

Topics

KU

Efficiency of OLS: The Gauss-Markov Theorem

Theorem: Gauss-Markov

Under Assumptions MLR.1 through MLR.5 (Gauss-Markov assumptions), the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1$, ..., $\hat{\beta}_k$ are the **best linear unbiased estimators (BLUEs)**

- To understand each component of the acronym "BLUE" let's start from the end.

Efficiency of OLS: The Gauss-Markov Theorem

E (estimator): It is a rule that can be applied to any sample of data to produce an estimate.

Efficiency of OLS: The Gauss-Markov Theorem

U (unbiased): $\hat{\beta}_j^{OLS}$ is an unbiased estimator of the true parameter, i.e., β_j.

$\Rightarrow E(\hat{\beta}_j^{OLS}) = \beta_j$ for any $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$

(conditional on $\{(x_{i1}, ..., x_{ik}) : i = 1, ..., n\}$).

Efficiency of OLS: The Gauss-Markov Theorem

L (linear): The estimator is a linear function of $\{y_i : i = 1, 2, ..., n\}$, but *it can be a nonlinear function of the explanatory variables.*, i.e.,

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$$

where the $\{w_{ij} : i = 1, ..., n\}$ are any functions of $\{(x_{i1}, ..., x_{ik}) : i = 1, ..., n\}$.

- The OLS estimators can be written in this way.

Efficiency of OLS: The Gauss-Markov Theorem

B (best): This means smallest variance (which makes sense once we impose unbiasedness).

$$Var(\hat{\beta}_j) \leq Var(\tilde{\beta}_j) \text{ all } j$$

usually the inequality is strict. (conditional on the explanatory variables in the sample).

• If we do not impose unbiasedness, then we can use silly rules – such as $\tilde{\beta}_j = 1$ always – to get estimators with zero variance.

Efficiency of OLS: The Gauss-Markov Theorem

- If the Gauss-Markov assumptions hold, and we insist on unbiased estimators that are also linear functions of $\{y_i : i = 1, 2, ..., n\}$, then

OLS delivers the smallest possible variances.

- We are not looking nonlinear functions of $\{y_i : i = 1, 2, ..., n\}$.

Efficiency of OLS: The Gauss-Markov Theorem

- **Remember:** Failure of MLR.5 does not cause bias in the $\hat{\beta}_j$, but it does have two consequences:

1. The usual formuals for $Var(\hat{\beta}_j)$, and therefore for $se(\hat{\beta}_j)$, are wrong.

2. The $\hat{\beta}_j$ are no longer BLUE.