MULTIPLE REGRESSION ANALYSIS: ESTIMATION

Hüseyin Taştan[1]

[1]Yıldız Technical University
Department of Economics

These presentation notes are based on
Introductory Econometrics: A Modern Approach (2nd ed.)
by J. Wooldridge.

17 Ekim 2012

Multiple Regression Analysis

- In the simple regression analysis with only one explanatory variable the assumption SLR.3 is generally very restrictive and unrealistic.
- Recall that the key assumption was SLR.3: all other factors affecting y are uncorrelated with x (ceteris paribus).
- Multiple regression analysis allow us to explicitly control for many other factors that simultaneously affect y.
- By adding new explanatory variables we can explain more of the variation in y. In other words, we can develop more successful models.
- Additional advantage: multiple regression can incorporate general functional form relationships.

Multiple Regression Analysis Examples

The Model with Two Explanatory Variables

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

Wage Equation

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

$wage$: hourly wages (in US dollars); $educ$: level of education (in years); $exper$: level of experience (in years)

- β_1: measures the impact of eduction on wage, holding all other factors fixed.
- β_2: measures the ceteris paribus effect of experience on wage.
- This wage equation allows us to measure the impact of education on wage holding experience fixed. This was not possible in simple regression analysis.

Multiple Regression Analysis Examples

Student Success and Family Income

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

$avgscore$: average standardized test score; $expend$: education spending per student; $avgincome$: average family income

- If $avginc$ is not included in the model directly, its effect will be included in the error term, u.
- Because average family income is generally correlated with education expenditures, the key assumption SLR.3 will be invalid: x ($expend$) will be correlated with u leading to biased OLS estimators.

Multiple Regression Analysis

- Multiple regression analysis allow us to use more general functional forms.
- Consider the following quadratic model of consumption:

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

where $x_1 = inc$ and $x_2 = inc^2$

- β_1: we cannot fix inc^2 while changing inc.
- **Marginal propensity to consume** (MPC) is approximated by:

$$\frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc$$

- MPC depends on the level of income.

Multiple Regression Analysis

The Model with Two Explanatory Variables

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

- "Zero conditional mean" assumption:

$$\mathrm{E}(u|x_1, x_2) = 0$$

- In other words, for all combinations of x_1 and x_2, the expected value of unobservables, u, is zero.
- E.g., in the wage equation:

$$E(u|educ, exper) = 0$$

- This means that unobservable factors affecting wage are not related on average with $educ$ and $exper$.
- If ability is a part of u, then average ability levels must be the same across all combinations of education and experience in the population.

Multiple Regression Analysis

The model with two independent variables

$$\mathrm{E}(u|x_1, x_2) = 0$$

- Test scores and average family income:

$$E(u|expend, avginc) = 0$$

- All other factors affecting the average test scores (such as quality of schools, student characteristics, etc.) are, on average, unrelated to education expenditures and family income.
- In the quadratic consumption model:

$$E(u|inc, inc^2) = E(u|inc) = 0$$

- Since inc^2 is automatically known when inc is known we do not need to write it in the conditional expectation.

The Model with k Independent Variables

Model

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

- In this model we have k explanatory variables and $k + 1$ unknown β parameters.
- The definition of the error term is the same: it represents all other factors affecting y that are not included in the model.

The Model with k Explanatory Variables

- β_j: measures the change in y in response to a unit change in x_j holding all other xs and unobservable factors u fixed.
- For example:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 ceoten + \beta_3 ceoten^2 + u$$

$ceoten$: tenure with the current company (years)

- Interpretation of β_1: elasticity of salary with respect to sales. Ceteris paribus salary will change by β_1% in response to a 1% change in sales.
- β_2: does not measure the impact of $ceoten$. We need to take the quadratic term into account.

The Model with k Explanatory Variables

Zero Conditional Mean Assumption

$$\mathrm{E}(u|x_1, x_2, \ldots, x_k) = 0$$

- Error term is unrelated with explanatory variables.
- If u is correlated with any of xs then OLS estimators will in general be biased. Estimation results will be unreliable.
- If there are omitted important variables affecting y this assumption may not hold. This may lead to "omitted variable bias".
- This assumption also means that functional form is correctly specified.

The Model with k Explanatory Variables: OLS Estimation

Sample Regression Function - SRF

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

Ordinary Least Squares (OLS) estimators minimize the Sum of Squared Residuals (SSR):

OLS Objective Function

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

OLS estimators, $\hat{\beta}_j$, $j = 0, 1, 2, \ldots, k$, can be found by solving the First Order Conditions (FOC) involving $k+1$ equations with $k+1$ unknowns.

The Model with k Explanatory Variables: OLS Estimation

OLS First Order Conditions

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\sum_{i=1}^{n} x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik}) &= 0 \\
\vdots \quad \vdots \quad \vdots & \\
\sum_{i=1}^{n} x_{ik} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik}) &= 0
\end{aligned}$$

The Model with k Explanatory Variables: OLS Estimation

OLS and Method of Moments

OLS first order conditions can also be obtained using the method of moments. They are the sample counterparts to the following population moment conditions:

$$\begin{aligned} \mathsf{E}(u) &= 0 \\ \mathsf{E}(x_1 u) &= 0 \\ \mathsf{E}(x_2 u) &= 0 \\ \vdots &\ \vdots\ \vdots \\ \mathsf{E}(x_k u) &= 0 \end{aligned}$$

Interpretation of SRF

The model with 2 explanatory variables

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

- Slope parameter estimates, $\hat{\beta}_j$, gives us the predicted change in y given the changes in x variables, ceteris paribus.
- $\hat{\beta}_1$: holding x_2 fixed, i.e. $\Delta x_2 = 0$

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1$$

- Similarly, holding x_1 fixed

$$\Delta \hat{y} = \hat{\beta}_2 \Delta x_2$$

Example: Determinants of College Success, gpa1.gdt

$colGPA$ Estimation Results

$$\widehat{colGPA} = 1.29 + 0.453\ hsGPA + 0.0094\ ACT$$

$n = 141$ students, $colGPA$: university grade point average (GPA, points out of 4), $hsGPA$: high school grade point average, ACT: achievement test score

- Intercept term is estimated as $\hat{\beta}_0 = 1.29$. This gives us the average $colGPA$ when $hsGPA = 0$ and $ACT = 0$. Because there is no observation with $ACT = 0$ and $hsGPA = 0$, this interpretation is not meaningful.
- Slope parameter on $hsGPA$ is estimated to be 0.453. Holding ACT fixed, given a one-point change in high school GPA, college GPA is predicted to increase by 0.453 points, almost half a point.

Example: Determinants of College Success, gpa1.gdt

$colGPA$ Estimation Results

$$\widehat{colGPA} = 1.29 + 0.453\ hsGPA + 0.0094\ ACT$$

$n = 141$ students, $colGPA$: university grade point average (GPA, points out of 4), $hsGPA$: high school grade point average, ACT: achievement test score

- Alternative interpretation of 0.453
- If we choose two students, A and B, with the same ACT score but $hsGPA$ of student A is one point higher than the $hsGPA$ of student B, then we predict that student A's college GPA is 0.453 points higher than that of student B's.
- ACT has $+$ sign but its effect is very small.

Example: Determinants of College Success, gpa1.gdt

- Regression with only ACT variable:

$colGPA$ simple regression results

$$\widehat{colGPA} = 2.4 + 0.0271\ ACT$$

- The coefficient estimate on ACT is almost three times as large as the previous estimate.
- But this regression does not allow us to compare two students with the same high school GPA
- The impact ACT decreases considerably after controlling for the high school GPA.

Interpretation of Estimated Regression Function

SRF - Sample Regression Function

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

In terms of changes

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

- Interpretation of the coefficient on x_1, $\hat{\beta}_1$: holding all other variables fixed, i.e. $\Delta x_2 = 0$, $\Delta x_3 = 0$, ..., $\Delta x_k = 0$

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1$$

- Holding all other independent variables fixed (i.e. controlling for all other x variables), the change in $\hat{y}$ given a one unit change in x_1 is $\hat{\beta}_1$ (in terms of y's unit of measurement)

Example: Logarithmic Wage Equation

Estimated equation

$$\widehat{log(wage)} = 0.284 + 0.092educ + 0.0041exper + 0.022tenure$$

$n = 526$ workers.

- Functional form: log-level, $100\hat{\beta}_j$ gives us % change in wages.
- For example, holding experience and tenure fixed, another year of education is predicted to increase 9.2% change in wages.
- In other words, given two workers with the same level of experience and tenure, if one of the workers has a level of education one year higher than the other one, then the predicted difference between their wages is 9.2

Changing More than One Explanatory Variable Simultaneously

- Sometimes we want to change more than one x variables at the same time to find their impact on y.
- Also in some cases when of the x variables changes the other automatically changes.
- For example, in the wage equation when we increase tenure 1 year experience also increases by 1 years
- In this case the total effect on wage is % 2.61:

$$\Delta\widehat{log\ Ucret} = 0.0041\Delta tecrube + 0.022\Delta kidem$$

$$= 0.0041 + 0.022 = 0.0261$$

OLS Fitted Values and Residuals

Fitted (predicted) value for observation i

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik}.$$

Residual for observation i

$$\hat{u}_i = y_i - \hat{y}_i$$

- $\hat{u} > 0$ implies $y_i > \hat{y}_i$, underprediction
- $\hat{u} < 0$ implies $y_i < \hat{y}_i$, overprediction

Algebraic Properties of Residuals

- The sum (and also average) of OLS residuals is zero:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \bar{\hat{u}} = 0$$

  follows directly from the first moment condition.
- Sample covariance between OLS residuals and each x_j is zero:

$$\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0, \quad j = 1, 2, \ldots, k$$

  also follows directly from the population moment conditions. Imposed by the OLS FOC.
- The point $(\bar{x_j}, \bar{y} : j = 1, 2, \ldots, k)$ is always on the OLS regression line.
- $\bar{y} = \bar{\hat{y}}$

An Alternative Derivation of Coefficient Estimates

The model with two explanatory variables

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

- The slope estimator on x_1 is:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$

- where $\hat{r}_{i1}$ is the residual obtained from the regression of x_1 on x_2:

$$x_{i1} = \hat{\alpha}_0 + \hat{\alpha}_1 x_{i2} + \hat{r}_{i1}$$

- This means that $\hat{\beta}_1$ is the slope estimate obtained from regressing y on residuals, $\hat{r}_{i1}$:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 r_{i1} + residual$$

- Ceteris paribus (partialling out, netting out): $\hat{\beta}_1$ measures the sample relationship between y and x_1 after x_2 has been partialled out.

Comparison of Simple and Multiple Regression Estimates

Two models:

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1, \quad vs. \quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

- These regressions will generally give different results.
- There are two special cases in which they give the same estimates:
- The partial effect of x_2 on y is zero, $\hat{\beta}_2 = 0$
- x_1 and x_2 are uncorrelated in the sample.

Goodness-of-Fit and Sums of Squares

- SST: total variation in y.

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

  Note that: Var$(y) = SST/(n-1)$.
- SSE is the variation in the explained part:

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

- SSR is the variation in the unexplained part:

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

- Thus the total variation in y can be written as:

$$SST = SSE + SSR$$

Goodness-of-fit

- Dividing both sides of this expression by SST:

$$1 = \frac{SSE}{SST} + \frac{SSR}{SST}$$

- Coefficient of determination: the ratio of the variation in the explained part to the total variation:

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

- Since SSE can never be larger than SST we have $0 \leq R^2 \leq 1$
- R^2 measures the percentage of the variation in y that can be explained by the variations in x variables. This can be used as goodness-of-fit measure.
- R^2 can also be calculated as follows: $R^2 = Corr(y, \hat{y})^2$

Goodness-of-fit

- Coefficient of Determination:

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

- When a new x variable is added to the regression R^2 always increases (or stays the same). It never decreases.
- The reason is that, when a new variable is added SSR always decreases.
- For this reason R^2 may not be a good indicator model selection.
- Instead of R^2 we will use adjusted R^2 for model selection.

Goodness-of-fit: Example

College GPA

$$\widehat{colGPA} = 1.29 + 0.453\ hsGPA + 0.0094\ ACT$$

$$n = 141 \quad R^2 = 0.176$$

- The coefficient of determination is 0.176.
- %17.6 of the total variation in college GPA can be explained by the variations in $hsGPA$ and ACT.
- This is not very large because there are many factors not included in the model.

Regression through the Origin

Predicted y is 0 when all x_js are 0

$$\tilde{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \ldots + \tilde{\beta}_k x_k.$$

- Note that the regression does not have a constant term.
- R^2 can be negative in this regression. In this case R^2 is treated as 0 or the regression is re-estimated after adding an intercept term.
- $R^2 < 0$ means that the sample average of y, $\bar{y}$, is more successful in explaining total variation than the model including x variables.
- When the constant term in the PRF is different from 0, then the OLS estimators in the regression through the origin will be biased.
- Adding a constant term, when in fact it is 0, on the other hand, increases the variance of OLS estimators.

Assumptions for Unbiasedness of OLS Estimators

MLR.1 Linear in Parameters

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

Population model is linear in parameters.

MLR.2 Random Sampling

We have a random sample of n observations drawn from the population model defined in MLR.1:

$$\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : \ i = 1, 2, \ldots, n\}$$

Assumptions for Unbiasedness of OLS Estimators

MLR.3 Zero Conditional Mean

$$\mathrm{E}(u|x_1, x_2, \ldots, x_k) = 0$$

- This assumption states that x explanatory variables are **strictly exogenous**. Random error term is uncorrelated with explanatory variables.
- There are several ways this assumption can fail.
- 1. Functional form misspecification: if the functional relationship between the explained and explanatory variables is misspecified this assumption can fail.
- 2. Omitting an important variable that is correlated with any of explanatory variables.
- 3. Measurement error in explanatory variables.
- If this assumption fails then we have **endogenous** explanatory variables.

Assumptions for Unbiasedness of OLS Estimators

MLR.4 No Perfect Collinearity

There is no perfect linear relationship between all independent x variables. Any of the x variables cannot be written as a linear combination of other independent variables. If this assumption fails we have perfect collinearity.

- If x variables are perfectly collinear then it is not possible to mathematically determine OLS estimators.
- This assumption allows the independent variables to be correlated: but they cannot be perfectly correlated.
- If we did not allow for any correlation among the independent variables, then multiple regression would be of very limited use for econometric analysis.
- For example, in the regression of student GPA on education expenditures and family income, we suspect that expenditure and income may be correlated and so we would like to hold income fixed to find the impact of expenditure on GPA.

Finite Sample Properties of OLS Estimators

THEOREM: Unbiasedness of OLS

Under Assumptions MLR.1 through MLR.4 OLS estimators are unbiased:

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

The centers of the sampling distributions (i.e. expectations) of OLS estimators are equal to the unknown population parameters. Remember that an estimate cannot be unbiased, it is just a number based on a sample, and in general we cannot say anything about its proximity to the true value.

Including Irrelevant Variables in a Regression Model

- What happens if we add an irrelevant variable in the model? (overspecifying the model)
- Irrelevance of the variable means that its coefficient in the population model is zero.
- E.g., suppose that in the regression below the partial effect of x_3 is zero, $\beta_3 = 0$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- Taking the conditional expectation we have:

$$\mathsf{E}(y|x_1, x_2, x_3) = \mathsf{E}(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

- Even though the true model is given above x_3 is added to the model by mistake.

Including Irrelevant Variables in a Regression Model

- In this case SRF is given by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

- OLS estimators are still unbiased:

$$\mathsf{E}(\hat{\beta}_0) = \beta_0, \quad \mathsf{E}(\hat{\beta}_1) = \beta_1, \quad \mathsf{E}(\hat{\beta}_2) = \beta_2, \quad \mathsf{E}(\hat{\beta}_3) = 0,$$

- True parameter value for the irrelevant variable is 0. Since this variable would have no explanatory power the expected value of the OLS estimator will also be zero - i.e. unbiased.
- However, even if they are unbiased, the variance of the regression will be larger if the model is overspecified.

Omitting a Relevant Variable

- What happens if we exclude an important variable?
- If a relevant variable is omitted this implies that its parameter is **not** 0 in the PRF. This is called underspecification of the model.
- In this case OLS estimators will be **biased**.
- E.g. suppose that the PRF includes 2 independent variables:

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

- Suppose that we omitted x_2 because, say, it is unobservable. Now the SRF is

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

- Is the OLS estimator on x_1, $\tilde{\beta}_1$, still unbiased?

Omitting a Relevant Variable

- The impact of the omitted variable will be included in the error term:
$$y = \beta_0 + \beta_1 x_1 + \nu$$
- True PRF includes x_2'yi. Thus the error term ν can be written as:
$$\nu = \beta_2 x_2 + u$$
- OLS estimator of β_1 in the model above is:
$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$
- To determine the magnitude and sign of the bias we will substitute y in the formula for $\tilde{\beta}_1$, re-arrange and take expectation.

Omitting a Relevant Variable

$$\begin{aligned}\tilde{\beta}_1 &= \frac{\sum(x_{i1} - \bar{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i)}{\sum(x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} + \frac{\sum(x_{i1} - \bar{x}_1)u_i}{\sum(x_{i1} - \bar{x}_1)^2}\end{aligned}$$

Taking (conditional) expectation we obtain

$$\begin{aligned}E(\tilde{\beta}_1) &= \beta_1 + \beta_2 \frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} + \frac{\sum(x_{i1} - \bar{x}_1)\overbrace{\mathsf{E}(u_i)}^{=0, MLR.3}}{\sum(x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \beta_2 \left(\frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} \right)\end{aligned}$$

Omitting a Relevant Variable

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \left(\frac{\sum(x_{i1} - \bar{x}_1)x_{i2}}{\sum(x_{i1} - \bar{x}_1)^2} \right)$$

The expression in the parenthesis to the right of β_2 is just the regression of x_2 on x_1:

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$$

Thus

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1$$

$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$$

This is called **omitted variable bias**.

Omitted Variable Bias

$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$$

- Iff $\tilde{\delta}_1 = 0$ or $\beta_2 = 0$ then bias is 0.
- The sign of bias depends on both β_2 and the correlation between omitted variable (x_2) and included variable (x_1).
- It is not possible to calculate this correlation if omitted variable cannot be observed.
- The following table summarizes possible cases:

Direction of Bias

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	positive bias	negative bias
$\beta_2 < 0$	negative bias	positive bias

Omitted Variable Bias

$$bias = E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$$

- The size of the bias is also important. It depends on both $\tilde{\delta}_1$ and β_2.
- A small bias relative to β_1 may not be a problem in practice.
- But in most cases we are not able to calculate the size of the bias.
- In some cases we may have an idea about the direction of bias. For example, suppose that in the wage equation true PRF contains both education and ability.
- Suppose also that ability is omitted because it cannot be observed, leading to omitted variable bias.
- In this case we can say that sign of the bias is $+$ because it is reasonable to think that people with more ability tend to have higher levels of education and ability is positively related to wage.

Omitted Variable Bias

- The effect of omitted variable will be in u. Thus, MLR. 3 fails.

$$wage = \beta_0 + \beta_1 educ + \beta_2 ability + u$$

- Instead of this model we estimate

$$wage = \beta_0 + \beta_1 educ + \nu$$

$$\nu = \beta_2 ability + u$$

- Education will be correlated with the error term (ν).

$$\mathrm{E}(\nu | educ) \neq 0$$

- Education is endogenous. If we omit ability the effect of education on wages will be overestimated. Some part of the effect of education on wage comes from ability.

Omitted Variable Bias

- In models with more than two variables, omitting a relevant variable generally causes OLS estimators to be biased.
- True PRF is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

- x_3 is omitted and the following model is estimated

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

- Also suppose that x_3 is correlated with x_1 but uncorrelated with x_2.
- In this case both $\tilde{\beta}_1$ and $\tilde{\beta}_2$ will be biased. Iff x_1 and x_2 are correlated then $\tilde{\beta}_2$ will be unbiased.

The Variance of the OLS Estimators

Assumption **MLR.5: Homoscedasticity**

This assumption states that conditional on x variables error term has constant variance:

$$\mathrm{Var}(u | x_1, x_2, \ldots, x_k) = \sigma^2$$

- If this assumption fails the model exhibits **heteroscedasticity**.
- This assumption is essential in deriving variances and standard errors of OLS estimators and in showing whether OLS estimators are efficient.
- We do not need this assumption for unbiasedness.
- E.g., in the wage equation this assumption implies that, the variance of unobserved factors does not change with the factors included in the model (education, experience, tenure, etc.).

The Variance of the OLS Estimators

Gauss-Markov Assumptions

Assumptions MLR.1 through MLR.5 are called *Gauss-Markov assumptions*. **MLR.1**: Linear in parameters,
MLR.2: Random sampling,
MLR.3: Zero conditional mean,
MLR.4: No perfect collinearity,
MLR.5: Homoscedasticity.

- This assumptions are only valid for cross-sectional data. They need to be modified for time series data.
- Assumptions MLR.3 and MLR.5 can be restated in terms of the dependent variable:

$$\mathsf{E}(y|x_1, x_2, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$
$$\mathsf{Var}(y|x_1, x_2, \ldots, x_k) = \mathsf{Var}(u|x_1, x_2, \ldots, x_k) = \sigma^2$$

The Variance of the OLS Estimators

Theorem: Variances of $\hat{\beta}$

Under Gauss-Markov assumptions (MLR.1-MLR.5):

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

where

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

is the total sample variation in x_j and R_j^2 is the R-squared from regressing x_j on all other independent variables (including an intercept term).

- $\mathsf{Var}(\hat{\beta}_j)$ moves in the same direction with σ_2 but in opposite direction with SST_j.
- To increase SST_j we need to collect more data (increase n). To reduce σ^2 we need to find good explanatory variables.

The Variance of the OLS Estimators

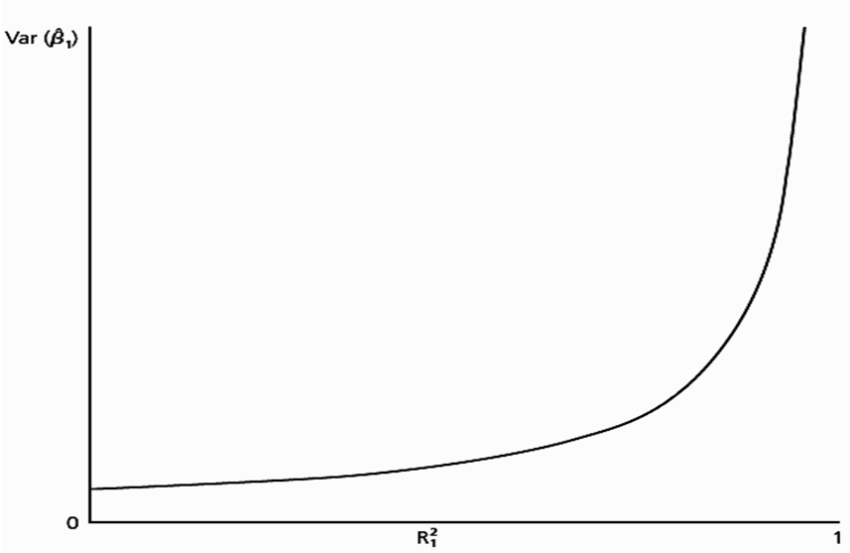
Theorem: Variances of $\hat{\beta}$

Under Gauss-Markov assumptions (MLR.1-MLR.5):

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

- $\mathsf{Var}(\hat{\beta}_j)$ also depends on R_j^2 which measures the degree of correlation among x variables.
- We did not have this term in the simple regression analysis because there was only one explanatory variable.
- As the degree of correlation increases among x variables the variances of OLS estimators get larger and larger.
- When there are high level of collinearity among x variables variances of OLS estimators will be larger. This is called multicollinearity problem.
- In the limit $R_j^2 = 1$ in which case the variance is infinite (note that in this case $\hat{\beta}_j$s will be indefinite. But MLR.4 prohibits

Relationship between variance and R_j^2



The Variance of the OLS Estimators

Estimating Variances

An unbiased estimator of the error variance is:

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

- Degrees of freedom:

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

- dof = number of observations − number of parameters
- dof comes from the OLS first order conditions. Recall that there were $k+1$ conditions. In other words $k+1$ restrictions are imposed on residuals.
- This means that given $n-(k+1)$ of the residuals, the remaining $k+1$ residuals are known: there are only $n-k-1$ degrees of freedom in the residuals.
- Notice that the error term u has n degrees of freedom.

The Variance of the OLS Estimators

Standard **deviations** of $\hat{\beta}_j$s

$$\text{sd}(\hat{\beta}_j) = \frac{\sigma}{\sqrt{SST_j(1-R_j^2)}}, \quad j = 1, 2, \ldots, k$$

Standard **errors** of $\hat{\beta}_j$s

$$\text{se}(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}, \quad j = 1, 2, \ldots, k$$

- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is the **standard error of regression** (SER).
- SER is an estimator of the standard deviation of the error term. SER may increase or decrease when a new variable is added in the model.
- $\text{se}(\hat{\beta}_j)$ is used in calculating confidence intervals and testing hypotheses.

Efficiency of OLS

Gauss-Markov Theorem

Under the assumptions MLR.1 through MLR.5, OLS estimators are the best linear unbiased estimators (BLUE) for unknown population parameters. In other words, under MLR.1-MLR.5, OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$ are BLUEs of $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$

- The Gauss-Markov theorem provides justification for the OLS estimation.
- If this theorem is valid then we do not need to find other estimation methods than OLS. The method of OLS gives us the most efficient (best, minimum variance) estimators.
- If any of the five assumptions fails the Gauss-Markov theorem does not hold. When MLR.3 fails unbiasedness property does not hold. When MLR.5 fails efficiency does not hold.

The Meaning of Linearity of Estimators in BLUE

Linearity of Estimators

An estimator $\tilde{\beta}_j$ of β_j is linear, iff, it can be expressed as a linear function of the data on the dependent variable:

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

where each w_{ij} can be a function of the sample values of all the independent variables.

The OLS estimators can be written in the form given above:

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n}\hat{r}_{ij}y_i}{\sum_{i=1}^{n}\hat{r}_{ij}^2} = \sum_{i=1}^{n} w_{ij} y_i, \quad \text{where} \quad w_{ij} = \frac{\hat{r}_{ij}}{\sum_{i=1}^{n}\hat{r}_{ij}^2}$$

where $\hat{r}_{ij}$ is the residual obtained from the regression of x_j on all other explanatory variables.