THE MULTIPLE REGRESSION MODEL II

PROF. SEBASTIÁN VALLEJO VERA

GOODNESS-OF-FIT

As with the simple regression model, we have that:

1. $TSS \equiv \sum_{i=1}^{n}(y_i - \bar{y})^2$
   - *TSS* is the total sum of squares, and it measures the total sample variation in the y_i; that is, it measures how spread out the y_i are in the sample, or how much there is to explain.
2. $ESS \equiv \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$
   - *ESS* is the explained sum of squares and, similarly, measures the sample variation in the $\hat{y}_i$; that is, how much of the variation is explained by the model.
3. $RSS \equiv \sum_{i=1}^{n} \hat{\mu}_i^{\,2}$

The total variation in y can be expressed as the sum of the explained variation and the unexplained variation: $TSS = ESS + RSS$.

- The R-squared of the regression is similarly defined and interpreted:

$$R^2 \equiv ESS/TSS = 1 - RSS/TSS$$

- An important fact about R^2 is that it never decreases, and it usually increases when another independent variable is added to a regression.
- The fact that R^2 never decreases when any variable is added to a regression makes it a poor tool for deciding whether one variable or several variables should be added to a model. The factor that should determine whether an explanatory variable belongs in a model is whether the explanatory variable has a nonzero partial effect on y in the population.
- Again, a low R^2 should not be source for worry because it generally indicates that the phenomenon at hand, like human behaviour, is hard to explain.

- As with the simple regression model, we can study the properties of the distributions of $\widehat{\beta_0}$, $\widehat{\beta_1}$, …, $\widehat{\beta_k}$ over repeated random samples from the population.

- Let’s use “MLR” to refer to Multiple Linear Regression.

THE ASSUMPTIONS

THAT LEAD TO THE GAUSS-MARKOV THEOREOM

ASSUMPTION MLR.1: LINEAR IN PARAMETERS

The model in the population can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \mu$$

where $\beta_0, \beta_1, \beta_2, \beta_3, \ldots, \beta_k$ are the unknown parameters of interest and μ is an unobserved random error or disturbance term.

ASSUMPTION MLR.2: RANDOM SAMPLING

We have a random sample of size n, $\{(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ik}, y_i): i = 1, \ldots, n\}$, following the population model:

$$y = \beta_0 + \ \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \mu$$

ASSUMPTION MLR.3: NO PERFECT COLLINEARITY

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

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

It is important to note that Assumption MLR.3 does allow the independent variables to be correlated; they just cannot be *perfectly* correlated.

- The simplest way that two independent variables can be perfectly correlated is when one variable is a constant multiple of another.
- Another way that independent variables can be perfectly collinear is when one independent variable can be expressed as an exact linear function of two or more of the other independent variables.
- Finally, Assumption MLR.3 also fails if the sample size, n, is too small in relation to the number of parameters being estimated. In the general regression model, there are $k + 1$ parameters, and MLR.3 fails if $n < k + 1$. Intuitively, this makes sense: to estimate $k + 1$ parameters, we need at least $k + 1$ observations.

ASSUMPTION MLR.4: ZERO CONDITIONAL MEAN

The error μ has an expected value of zero given any value of the explanatory variable x. In other words:

$$E(\mu|x_1, x_2, x_3, \ldots, x_k) = 0$$

When Assumption MLR.4 holds, we often say that we have **exogenous explanatory variables**. If x_j is correlated with μ for any reason, then x_j is said to be an **endogenous explanatory variable**.

- Omitting a determinant of y that is correlated with any of $x_1, x_2, x_3, \dots, x_k$ causes Assumption MLR.4 to fail.
- Another way that Assumption MLR.4 can fail is if the functional form (relationship) between the *explained* and *explanatory* variables is misspecified.
- There are other instances when Assumption MLR.4 fails (e.g., measurement error in the independent variable(s), when one or more of the independent variables are jointly determined with y , etc.) but these issues will not be treated here.

THE GAUSS-MARKOV THEOREM

THEOREM MLR.1: UNBIASEDNESS OF OLS

Under assumptions MLR.1 through MLR.4,

$$E\left(\widehat{\beta}_j\right) = \beta_j, j = 0,1, \dots, k$$

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

Remember that an estimate cannot be unbiased: an estimate is a fixed number, obtained from a particular sample, which usually is not equal to the population parameter.

When we say that OLS is unbiased under Assumptions MLR.1 through MLR.4, we mean that the *procedure* by which the OLS estimates are obtained is unbiased when we view the procedure as being applied across all possible random samples.

- Suppose that we omit a variable that actually belongs in the true (or population) model. This is often called the problem of excluding a **relevant variable, omitted variable,** or **underspecifying the model** and generally causes the OLS estimators to be biased.

- Suppose the *true* population model has two explanatory variables and an error term:

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

- and we assume that this model satisfies Assumptions MLR.1 through MLR.4.

- But suppose we don't have a measure of x_2 and estimate instead the underspecified equation:

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

He can demonstrate (though not here) that $\tilde{\beta}_1$ is equal to:

$$\tilde{\beta}_1 = \widehat{\beta_1} + \widehat{\beta_2}\widetilde{\delta_1}$$

Where $\widehat{\beta_1}$ and $\widehat{\beta_2}$ com from:

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2$$

And $\widetilde{\delta_1}$ from:

$$\widehat{x_2} = \widetilde{\delta_0} + \widetilde{\delta_1}x_1$$

The bias from omitting x_2 under the Assumptions MLR.1 through MLR.4 (where $\widehat{\beta_2}$ would be unbiased for β_2) is:

$$Bias(\tilde{\beta}_1) = \beta_2\widetilde{\delta_1}$$

From $Bias(\tilde{\beta}_1) = \beta_2 \widetilde{\delta_1}$, we see that there are two cases where $\tilde{\beta}_1$ is unbiased:

- when $\beta_2 = 0$, that is, when x_2 does not appear in the true population model
- when $\widetilde{\delta_1} = 0$, that is, when x_1 and x_2 are uncorrelated

When x_1 and x_2 are correlated and $\beta_2 \neq 0$, $\tilde{\beta}_1$ is biased in repeated samples (note that such statement does not hold in one sample).

The sign of the bias in $\tilde{\beta}_1$ depends on the signs of both β_2 and $\widetilde{\delta_1}$ (knowing that when x_1 and x_2 are positively correlated $\widetilde{\delta_1} > 0$ and when they are negatively correlated $\widetilde{\delta_1} < 0$).

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ when x_2 Is Omitted in Estimating Eqution (3.40)		
	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

- In practice, since β_2 is an unknown population parameter, we cannot be certain whether β_2 is positive or negative, although we usually have a pretty good idea about the direction of the partial effect of x_2 on y .
- Similarly, even though the sign of the correlation between x_1 and x_2 cannot be known if x_2 is not observed, in many cases, we can make an educated guess about whether x_1 and x_2 are positively or negatively correlated.

- Beyond the sign of the bias in $\tilde{\beta}_1$, we should also be concerned about the size of that bias.
- The size of the bias in $\tilde{\beta}_1$ also depends on the size of β_2 and $\widetilde{\delta_1}$. For example, if the true effect of x_2 on y is small (that is, β_2 is small) and/or x_2 on x_1 are only weakly correlated (that is, $\widetilde{\delta_1}$ is close to zero), then the bias in $\tilde{\beta}_1$ will be small. This is important because a large (small) bias should (not) be of great concern.
- In a model with k explanatory variables, guessing the sign and size of the bias introduced by the omission of a relevant variable is a much harder (if not impossible) task.

THEOREM MLR.2: SAMPLING VARIANCES OF THE OLS SLOPE ESTIMATORS

- Again, to facilitate the calculation of the variance of the OLS estimator, we state an additional assumption, this time about the variance of the unobservable, u, conditional on the explanatory variables.

Assumption MLR.5: Homoskedasticity

- The error u has the same variance given any values of the explanatory variables. In other words:

$$Var(\mu|x_1, x_2, x_3, \dots, x_k) = \sigma^2$$

- This fifth assumption is important because—as we will see—it provides OLS with an important efficiency property. Also, note that assumptions MLR.1 through MLR.5 are known as the Gauss-Markov assumptions (for cross-sectional data analysis).

- Under assumptions MLR.1 through MLR.5, conditional on the sample values of the independent variables,

$$Var(\widehat{\beta_j}) = \frac{\sigma^2}{TSS_j\left(1 - R_j^2\right)}$$

- for $j = 1, 2, \ldots, k$, where $\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2$ is the total sample variation in x_j, and R_j^2 is the R-squared from regressing x_j on all the other independent variables (and including an intercept).
- The size of $Var(\widehat{\beta_j})$ is important because a larger variance means a less precise estimator, and this translates into larger confidence intervals and less accurate hypotheses tests.

- The formula presented in Theorem MLR.2 tells us that the variance of $\widehat{\beta}_j$ depends on three factors: σ^2, TSS_j, and R_j^2.

1. **The error variance, $\boldsymbol{\sigma^2}$**: a larger σ^2 means larger variances for the OLS estimators. This is not at all surprising: more "noise" in the regression equation makes it more difficult to estimate the partial effect of any of the independent variables on y , and this is reflected in higher variances for the OLS slope estimators.

For a given dependent variable y, there is really only one way to reduce the error variance, and that is to add more explanatory variables to the equation (that is, take some factors out of the error term). Unfortunately, it is not always possible to find additional legitimate factors that affect y.

2. The total sample variation in x_j, TSS_j: the larger the total variation in x_j is, the smaller is $Var(\widehat{\beta_j})$. Thus, everything else being equal, for estimating β_j we prefer to have as much sample variation in x_j as possible.

Although it is rarely possible for us to choose the sample (or population) values of the independent variables, the sample variation in each of the independent variable increases with sample size. Thus, when possible, aim for a large sample size to gain in precision.

3. The Linear relationships among the independent variables, $\boldsymbol{R_j^2}$: R_j^2 is distinct from the R-squared in the regression of y on $x_1, x_2, x_3, \ldots, x_k$. R_j^2 is obtained from a regression involving only the independent variables in the original model, where xj plays the role of a dependent variable as in:

$$x_j = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \ \ldots + \ \gamma_{j-1} x_{j-1} + \ \gamma(j+1) x_{j+1} + \ \ldots + \ v$$

Because the R-squared measures goodness-of-fit, a value of R_j^2 close to one indicates that $x_1, x_2, x_3, \ldots, x_{j-1}, x_{j+1}, x_k$ explain much of the variation in x_j in the sample. In other words, it means that the proportion of the total variation in x_j , (TSS_j), is explained by the other independent variables $(x_1, x_2, x_3, \ldots, x_{j-1}, x_{j+1}, x_k)$.

It is easy to see that as R_j^2 gets closer to one, $Var(\widehat{\beta_j})$ gets larger and larger. High (but not perfect) correlation between two or more independent variables is called **multicollinearity**.

The problem, again, with $Var(\widehat{\beta_j}) = \frac{\sigma^2}{TSS_j\left(1-R_j^2\right)}$ is that we do not know the error variance, σ^2.

We can estimate it as follows:

$$\sigma^2 = \frac{\sum_{i=1}^{n} \widehat{\mu_i}^2}{(n-k-1)} = \frac{RSS}{(n-k-1)}$$

where $n - k - 1$ is **the degrees of freedom (df)** for the OLS estimation with n observations and k independent variables. Since there are $k + 1$ parameters in a regression model with k independent variables and an intercept, we can write:

$df = n - (k + 1)$ = (number of observations)–(number of observed parameters)

- This takes us to the following theorem…

THEOREM MLR.3: UNBIASED ESTIMATION OF σ^2

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

$$E\left(\widehat{\sigma^2}\right) = \sigma^2$$

- For constructing confidence intervals (CI) and conducting hypothesis tests, we will need the **standard deviation** of $\widehat{\beta}_j$:

$$sd(\widehat{\beta_j}) = \frac{\sigma}{[TSS_j(1 - R_j^2)]^{\frac{1}{2}}}$$

- But since σ is unknown, we replace it with its estimator, $\widehat{\sigma}$. This gives us the **standard error** of $\widehat{\beta}_j$:

$$se(\widehat{\beta_j}) = \frac{\widehat{\sigma}}{[TSS_j(1 - R_j^2)]^{\frac{1}{2}}}$$

- $se(\widehat{\beta_j})$ is what appears in the OLS regression outputs from most statistical softwares.

THEOREM MLR.4: GAUSS-MARKOV THEOREM

- Under assumptions MLR.1 through MLR.5, $\widehat{\beta_0}$, $\widehat{\beta_1}$, …, $\widehat{\beta_k}$ are the **best linear unbiased estimators (BLUEs)** of $\beta_0, \beta_1, \beta_2, \beta_3, \dots, \beta_k$, respectively.
- In other words, in the class of linear unbiased estimators, OLS has the smallest variance (under the five Gauss-Markov assumptions).
- *Best* in **BLUE** means having the *smallest* variance. Given two unbiased estimators, it is logical to prefer the one with the smallest variance because it means that the estimator is more precise, on average.

- Ok, that was… something… but how does it help me with life?