

Econometrics Review

Erick F Molina

Summer, 2024

ITAM

OLS Review

Asymptotic properties of OLS estimators

Using asymptotic properties, it is possible to show that:

$$\begin{aligned}\hat{\beta} &\xrightarrow{p} \beta \\ \sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} N(0, \alpha \Sigma \alpha')\end{aligned}\tag{1}$$

where

$$\alpha = E(X_i X_i')^{-1}$$

$$\Sigma = E(U_i^2 X_i X_i')$$

Homoskedasticity and heteroskedasticity

Assumptions of homoskedasticity and heteroskedasticity

Up to now, our only assumptions in the regression model have been:

- (i) that our sample is random (i.i.d.),
- (ii) the linear relationship between Y_i and X_i .

The homoskedasticity assumption states that the conditional variance of the errors is constant (i.e., it does not change with the level of X).

- This assumption is restrictive; however, it is not necessary in order to conduct inference.

If instead we assume heteroskedasticity:

- It is not necessary to assume that the variance of the errors is constant.
- We can estimate the variance of the coefficients with the sample analog of $\alpha \Sigma \alpha'$.

The heteroskedasticity case

To obtain a consistent estimator of $\alpha \Sigma \alpha'$ it is enough to use:

$$\hat{\alpha} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \xrightarrow{P} E(X_i X_i')^{-1} = \alpha \quad (2)$$

$$\hat{\Sigma} = \left(\frac{1}{n} \sum_{i=1}^n \hat{U}_i^2 X_i X_i' \right)^{-1} \xrightarrow{P} E(U_i^2 X_i X_i')^{-1} = \Sigma \quad (3)$$

$$\hat{U}_i = Y_i - X_i' \hat{\beta} \quad (4)$$

Using WLLN properties, we will have that:

$$\hat{\alpha} \hat{\Sigma} \hat{\alpha}' \xrightarrow{P} \alpha \Sigma \alpha' \quad (5)$$

The homoskedasticity case

In the homoskedasticity case, we assume that the variance of the errors given X_i is constant:

$$\text{Var}(U_i|X_i) = E(U_i^2|X_i) = \sigma^2 \quad (6)$$

Applying this assumption to the second factor of the variance:

$$E(U_i^2 X_i X_i') = E(E(U_i^2|X_i) X_i X_i') = \sigma^2 E(X_i X_i') \quad (7)$$

Therefore, the variance $\alpha \Sigma \alpha'$ reduces to:

$$\alpha \Sigma \alpha' = E(X_i X_i')^{-1} \sigma^2 E(X_i X_i') E(X_i X_i')^{-1} = \sigma^2 E(X_i X_i')^{-1} \quad (8)$$

Since $E(U_i^2)$ is a consistent estimator of σ^2 , to obtain a consistent estimator of the variance we only need:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{U}_i^2 \quad (9)$$

$$\hat{\sigma}^2 \hat{\alpha} \xrightarrow{P} \sigma^2 E(X_i X_i')^{-1} \quad (10)$$

The Frisch-Waugh-Lovell (FWL) Theorem

The Frisch-Waugh-Lovell (FWL) Theorem

- The Frisch-Waugh-Lovell (FWL) Theorem is a key result in econometrics.
- It shows the equivalence between coefficients in a multiple regression and a series of partial regressions.
- It is useful for understanding how to control for certain variables while estimating others.

Statement of the Frisch-Waugh-Lovell Theorem

- The Frisch-Waugh-Lovell Theorem states that in a multiple regression:

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

- The coefficients for β_1 can be obtained in three steps:
 1. Run the regression of Y on X_2 and obtain the residuals $\hat{\epsilon}_Y = M_2 Y$.
 2. Run the regression of each column of X_1 on X_2 and obtain the residuals $M_2 X_1$.
 3. Run the regression of $\hat{\epsilon}_Y$ on $M_2 X_1$ to obtain $\hat{\beta}_1$.
- Formally:

$\hat{\beta}_1$ from $Y = X_1\beta_1 + X_2\beta_2 + \epsilon$ is equal to $\hat{\beta}_1$ from $M_2 Y = M_2 X_1 \beta_1 + \text{residuals}$

The Frisch-Waugh-Lovell (FWL) Theorem

- The Frisch-Waugh-Lovell Theorem shows how to separate the effects of different groups of variables in a multiple regression model.
- It allows us to isolate the impact of a set of independent variables ($\mathbf{X}_1$) on the dependent variable ($\mathbf{y}$), by first removing the influence of another set of control variables ($\mathbf{X}_2$).
- This "partialling out" implies that we can understand how the variables of interest affect the outcome, independently of the other variables included in the model.
- It is especially useful in analyses where there are many variables and we want to focus on the net effect of a specific subset.
- It provides a clear methodology for dealing with multicollinearity and for better understanding causal relationships in economic data.

Matrices needed to prove the theorem

We need two important types of matrices to understand the algebraic proof of FWL.

1. Projection matrix $P = X(X'X)^{-1}X'$

- This matrix, when applied to an outcome vector (y), produces a set of predicted values ($\hat{y}$).
- Decomposing this, note that: $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Py$.

2. Residual-maker matrix M - the annihilator matrix:

$$M = I - X(X'X)^{-1}X'.$$

- It is the complement of Py .
- Since $My = y - X(X'X)^{-1}X'y \equiv y - Py \equiv y - X\hat{\beta} \equiv \hat{\epsilon}$, the residuals from regressing Y on X .

Matrices needed to prove the theorem

- Given these definitions, note that M and P are complementary:

$$y = \hat{y} + \hat{\epsilon} \equiv Py + My$$

- Therefore,

$$Iy = Py + My$$

$$I = P + M.$$

- With these projection matrices, we can express the FWL claim (which we need to prove) as:

$$y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\epsilon}$$

$$M_1y = M_1X_2\hat{\beta}_2 + \hat{\epsilon}.$$

Proof of the FWL Theorem

- Suppose that: $Y = X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\epsilon}$.
- First, we can multiply both sides by the annihilator matrix of X_1 :
$$M_1 Y = M_1 X_1 \hat{\beta}_1 + M_1 X_2 \hat{\beta}_2 + M_1 \hat{\epsilon}$$
- which first simplifies to: $M_1 Y = M_1 X_2 \hat{\beta}_2 + M_1 \hat{\epsilon}$,
- because $M_1 X_1 \hat{\beta}_1 \equiv (M_1 X_1) \hat{\beta}_1 \equiv 0 \hat{\beta}_1 = 0$.
- Second, we can simplify this equation further because, by the properties of OLS regression, X_1 and ϵ are orthogonal.
- Therefore, the residuals of the residuals are the residuals:
$$M_1 Y = M_1 X_2 \hat{\beta}_2 + \hat{\epsilon}$$
- This completes the proof.

Interpretation of coefficients

In general, we are interested in one of the parameters (β_k) that we are estimating. The interpretation of the estimated coefficient in the case of a multivariate regression is that it indicates the change in the dependent variable Y from a marginal change in the variable X_1 , holding all other control variables ($X_2, \dots, X_K$) constant.

Simple case

In the interpretation of coefficients, it is important to take into account in which units the dependent variable (Y) and the control variable that changes marginally (X_1) are measured.

Suppose we estimate the following specification with OLS:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + U_i$$

Let us see the result of increasing the variable X_1 by one unit ($\Delta X_1 = 1$) while holding all the other variables constant:

$$Y'_i = \beta_0 + \beta_1 (X_{1i} + 1) + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + U_i$$

In this case, the change in the dependent variable will be:

$$\Delta Y_i = Y'_i - Y_i = \beta_1$$

Therefore, β_1 will represent the change in Y_i that results from increasing X_1 by one unit while holding all other variables constant. During class we will see practical examples.

Functional forms

Even though the linear form of the OLS model specification may seem very restrictive in principle, it is possible to transform the independent variables and thereby achieve a better approximation of the relationship between two variables.

Carrying out transformations makes it possible to obtain an estimate that is more accurate in statistical terms and more correct in theoretical terms.

To decide which type of transformation is most appropriate to use in each case, one should consider:

- According to theory, which function most faithfully represents the relationship between two variables.
- According to the empirical data, which function describes the relationship among the observed data more accurately.

Polynomial transformation

In the case of transformations used to generate different functional forms, it is important to consider that the interpretations of the coefficients change in important ways.

To begin, consider the case of a polynomial transformation of some variable of interest. Suppose we estimate the following regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{1i}^2 + \dots + \beta_K X_{Ki} + U_i$$

If in this case we want to see the change in the dependent variable Y from a marginal change in X_1 , it is important to note that the variable X_1 appears in more than one term in the estimation. Therefore, if we want to approximate the effect of a marginal increase in X_1 (i.e., $\Delta X_1 = 1$), we compute the derivative with respect to X_1 :

$$\frac{\partial Y_i}{\partial X_{1i}} = \beta_1 + 2\beta_2 X_{1i}$$

Polynomial transformation (cont.)

What the previous equation indicates is that now, because of the quadratic form, the effect of a marginal increase in X_1 on Y depends on the level of X_1 from which we start. Therefore, the effect will differ depending on the individual. In particular, the coefficient β_1 corresponds to the change in Y associated with a one-unit increase in X_1 conditional on the initial value of $X_1 = 0$. On the other hand, β_2 has an interpretation related to how this marginal effect changes as X_1 increases:

$$\frac{\partial^2 Y_i}{\partial X_{1i}^2} = 2\beta_2$$

Polynomial transformation (cont.)

In the previous equation we can see that the sign of β_2 determines the sign of $\frac{\partial^2 Y_i}{\partial X_{1i}^2}$. That is, β_2 tells us whether the relationship between Y_i and X_{1i} is concave or convex.

In other words, it tells us whether the change in Y from a marginal change in X_1 is increasing or decreasing.

In the specification we used a polynomial of degree 2. However, we can use a higher-degree polynomial in the specification. It should be noted, however, that such a change must be supported theoretically or with empirical evidence (shown in the data), since as the polynomial degree increases, the interpretation of any effect based on the coefficients becomes more complicated.

Logarithmic transformation

In the case of a logarithmic transformation, the coefficients change in their interpretation. A particular characteristic of logarithms is that they cause the variable to which the logarithm is applied to no longer use its units in the interpretation. This is very helpful in cases in which the units being used are not easy to interpret or are not universally known (e.g., grading scales in a specific country). Instead of using its units, changes in this variable will be expressed in percentage terms. This is useful in cases in which percentage changes are common (e.g., monetary variables).

Logarithmic transformation (cont.)

The logarithmic transformation can be applied both to the dependent variable and to controls in the estimation. Let us start by seeing how to interpret the coefficient of some control when the dependent variable is a logarithm:

$$\ln(Y_i) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_K X_{Ki} + U_i$$

To see how we would now interpret a coefficient, consider the case of β_1 . First we compute the derivative with respect to X_1 :

$$\beta_1 = \frac{\partial \ln(Y_i)}{\partial X_{1i}} = \frac{\partial Y_i / \partial X_{1i}}{Y_i}$$

Logarithmic transformation (cont.)

In this case, for a change in X_{1i} of one unit (i.e., $\Delta X_{1i} = 1$), we have that β_1 represents a change in Y_i equal to $\frac{\Delta Y_i}{Y_i}$. To express this as a rate (in percentage terms) we multiply both sides by 100 to obtain:

$$\Delta Y_i = 100\beta_1$$

Therefore, the correct interpretation of the coefficient β_1 is that a one-unit change in X_{1i} entails a percentage change of β_1 in Y_i .

Similarly, if we use logarithms on both sides:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_{1i}) + \dots + \beta_K X_{Ki} + U_i$$

In this case, β_1 would be interpreted as the elasticity of Y with respect to X_1 , that is, a percentage change in X_1 results in a percentage change in Y .

Another interesting interpretation arises when the dependent variable is a dummy variable (i.e., a dichotomous variable $X_1 = \{0, 1\}$). In these To understand the interpretation of the coefficient of a dummy variable, it is convenient to review what the variable represents in the case of a simple linear regression. Suppose the dummy variable represents being a woman (i.e., $X_{1i} = 1$ if individual i is a woman and $X_{1i} = 0$ if the individual is a man).

Dummy variables

Suppose we estimate the following regression:

$$Calif_i = \beta_0 + \beta_1 X_{1i} + U_i$$

where $Calif_i$ represents the grade point average of person i . In this case, the conditional mean for being a woman and a man, respectively, are:

$$E(Calif_i \mid X_{1i} = 1) = \beta_0 + \beta_1$$

$$E(Calif_i \mid X_{1i} = 0) = \beta_0$$

Therefore, β_1 represents the difference in grades for being a woman relative to being a man:

$$\beta_1 = E(Calif_i \mid X_{1i} = 1) - E(Calif_i \mid X_{1i} = 0)$$

Interactions arise from using the product of two variables as an additional independent variable. In general, this product occurs between two dummy variables or between a dummy variable and a continuous one.

Interaction of two dummy variables

Consider a model where we want to explain total hours worked based on the individual's sex and marital status. Let X_{1i} be an indicator variable for being a woman ($X_{1i} = 1$ if i is a woman and $X_{1i} = 0$ if i is a man) and X_{2i} an indicator variable for marital status ($X_{2i} = 1$ if i is married and $X_{2i} = 0$ if i is single). Consider the following model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + U_i$$

Interactions (cont.)

To understand the interpretation of the interaction coefficient (β_3), consider conditional means again:

$$A = E(Y_i \mid X_{1i} = 0, X_{2i} = 0) = \beta_0$$

$$B = E(Y_i \mid X_{1i} = 0, X_{2i} = 1) = \beta_0 + \beta_2$$

$$C = E(Y_i \mid X_{1i} = 1, X_{2i} = 0) = \beta_0 + \beta_1$$

$$D = E(Y_i \mid X_{1i} = 1, X_{2i} = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

Interactions (cont.)

Therefore, we have:

$$\beta_2 = B - A$$

$$\beta_2 + \beta_3 = D - C$$

$$\beta_3 = (D - C) - (B - A)$$

β_2 represents the extra hours a married man works relative to a single man. $\beta_2 + \beta_3$ represents the extra hours a married woman works relative to a single woman. β_3 represents whether being married relative to being single implies a larger difference for a woman than for a man. Put differently, it indicates whether the “bonus” in hours worked from being married is different for a woman than for a man.

Interaction of a dummy variable and a continuous variable

Suppose now that one of our variables is continuous. Instead of the married/single dummy variable, suppose we have a variable that indicates the person's level of education. We might be interested in analyzing whether the returns to an additional year of education are similar for men and women. To do so we specify the following model:

$$\ln g_i = \beta_0 + \beta_1 \text{Educ}_i + \beta_2 \text{Mujer}_i + \beta_3 \text{Educ}_i \cdot \text{Mujer}_i$$

In this model $\ln g_i$ represents a person's monthly income measured in pesos; Educ_i the education level measured in completed years; and Mujer_i is a dummy variable equal to 1 if the person is a woman.

Dummy and continuous interaction

Using this model, the returns to an additional year of education for men are:

$$E(\ln g_i \mid \text{Mujer}_i = 0) = \beta_0 + \beta_1 \text{Educ}_i$$
$$\frac{\partial E(\ln g_i \mid \text{Mujer}_i = 0)}{\partial \text{Educ}_i} = \beta_1$$

Similarly, the returns to an additional year of education for women are:

$$E(\ln g_i \mid \text{Mujer}_i = 1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \text{Educ}_i$$
$$\frac{\partial E(\ln g_i \mid \text{Mujer}_i = 1)}{\partial \text{Educ}_i} = \beta_1 + \beta_3$$