EC 320: Introduction to Econometrics
Lecture Notes

David M. Hall
University of Oregon

Fall 2025

1 Session Overview

- Date:
- Topic:
  - Chapters (Dougherty):
- Goals/Objectives:
  1. One
  2. Two
  3. Three

1.1 Before Class:

2 During Class:

2.1 Welcome

- Introduce Multiple Regression Analysis
- Explain properties of MRA

2.1.1 Warm Up:

Example 2.1. No warm up today — lots to cover :)

2.2 Content Section #1:

Time:

Up until now, we have considered models where there is *one* explanatory variable s.t.

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

but there are not many phenomena — economic or not — that exhibit such behavior. The most classic example we use in econometrics, that I have strayed from just a bit, is a wage regression. I have strayed away from using this for the most part because I wanted to save wage talk for multiple regression specifically.

Ask students some determinants of wage.

Now, in the simple regression model we only would have tested *one* of those regressors on wage. For example, if you said "age" is a determinant of wage, then we would run:

$$WAGE_i = \beta_0 + \beta_1 AGE_i + u_i$$

In this regression, we leave out something like education, which is an unobserved variable in this regression and is therefore in u_i. The danger here, if you are testing what the effect of age is on wage, is conflating any of the age effect with other determinants due to not including education.

How do we explain why two 34-year-old individuals have totally different wages if we only include age? When one of them went to get their M.S. and one did not complete the third grade?

Or, what if we have a 34-year-old who got their M.S. and a 195 year-old who was never educated at all and works a minimum wage job. You might imagine in your head what the linear regression would do here.

What multiple regression allows us to do is estimate the effect of our regressors while *controlling* for the other variables. If done correctly, we will be able to separate out the effect of each regressor with a minimal set of assumptions.

2.3 Deriving the Multiple Regression Coefficient

We will very briefly practice with deriving the multiple regression coefficient. It follows the same steps as the single regression coefficient, just with now a larger system of equations. Our general case will utilize three regressors (including the intercept) as an illustrative exercise. I encourage anyone who is familiar with linear algebra to speak to me or consult an online source about deriving the OLS estimator using matrix algebra when we have many regressors.

We start by assuming that some variable, Y is deterimined by two explanatory variables and an intercept with the true model being describeed by:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \tag{1}$$

Note that we now have two subscripts for the X variables denoting which regressor it is. The u is still the disturbance term.

We will fit this line using the following:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} \tag{2}$$

How do we proceed?

Ask students to remind themselves what the LS in OLS stands for, and describe the steps.

Step 1: Isolate the residual.
Recall that our residual is defined as $\hat{u_i} = Y_i - \hat{Y}_i$. Thus our residual is:

$$\hat{u_i} = Y_i - \hat{\beta}_0 0 \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \tag{3}$$

We want to minimize the "sum of squares", so a natural next step might be to square our residual and apply the summation operator s.t.

$$RSS = \sum_i \hat{u_i^2} = \sum_i (Y_i - \hat{\beta}_0 0 \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})^2$$

I will note here that we can proceed one of two ways. I will do the second. The first way is to distribute the square and then derive. The second uses the chain rule from your calculus class.

Recall that for some differentiable function $u(x)$ that the chain rule says the derivative of this is $u'(x)dx$ where dx is the derivative of whatever "inner" function of x exists.

In our case, we are taking the derivative of our $\hat{\beta}$s where $\frac{d}{d\hat{\beta}}\hat{\beta}X = X$.

Partial w.r.t. $\hat{\beta}_0$:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = \sum_i 2\hat{u}_i \cdot (-1) = -2 \sum_i \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)$$

Partial w.r.t. $\hat{\beta}_1$:

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = \sum_i 2\hat{u}_i \cdot (-X_{1i}) = -2 \sum_i X_{1i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)$$

Partial w.r.t. $\hat{\beta}_2$:

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = \sum_i 2\hat{u}_i \cdot (-X_{2i}) = -2 \sum_i X_{2i} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} \right)$$

Step 4: Set each derivative to zero
To minimize the RSS, we set each of these partials equal to zero:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0, \quad \frac{\partial RSS}{\partial \hat{\beta}_1} = 0, \quad \frac{\partial RSS}{\partial \hat{\beta}_2} = 0$$

This yields the system of **normal equations**:

$$\sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_i X_{1i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

$$\sum_i X_{2i}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) = 0$$

Our next steps are to solve this system. Working with the first equation, we can get the equation for our intercept as a function of just the two explanatory variable means and coefficients:

$$\begin{aligned}
\sum_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i}) &= 0 \\
\sum_i Y_i - \sum_i \hat{\beta}_0 - \hat{\beta}_1 \sum_i X_{1i} - \hat{\beta}_2 \sum_i X_{2i} &= 0 \\
\sum_i Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_i X_{1i} - \hat{\beta}_2 \sum_i X_{2i} &= 0 \\
\text{Divide by n on both sides. Recall def'n. of mean.} & \\
\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 &= 0
\end{aligned}$$

$$\implies \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2 \tag{4}$$

Our next steps would be to substitute this into normal equation 2 to obtain $\hat{\beta}_1$ as a function of $\hat{\beta}_2$, and substitute it also into normal equation 3 along with the prior to obtain $\hat{\beta}_2$ as a function of just observed variables.

What you will get is... crazy enough for me not to want to write it all on here.

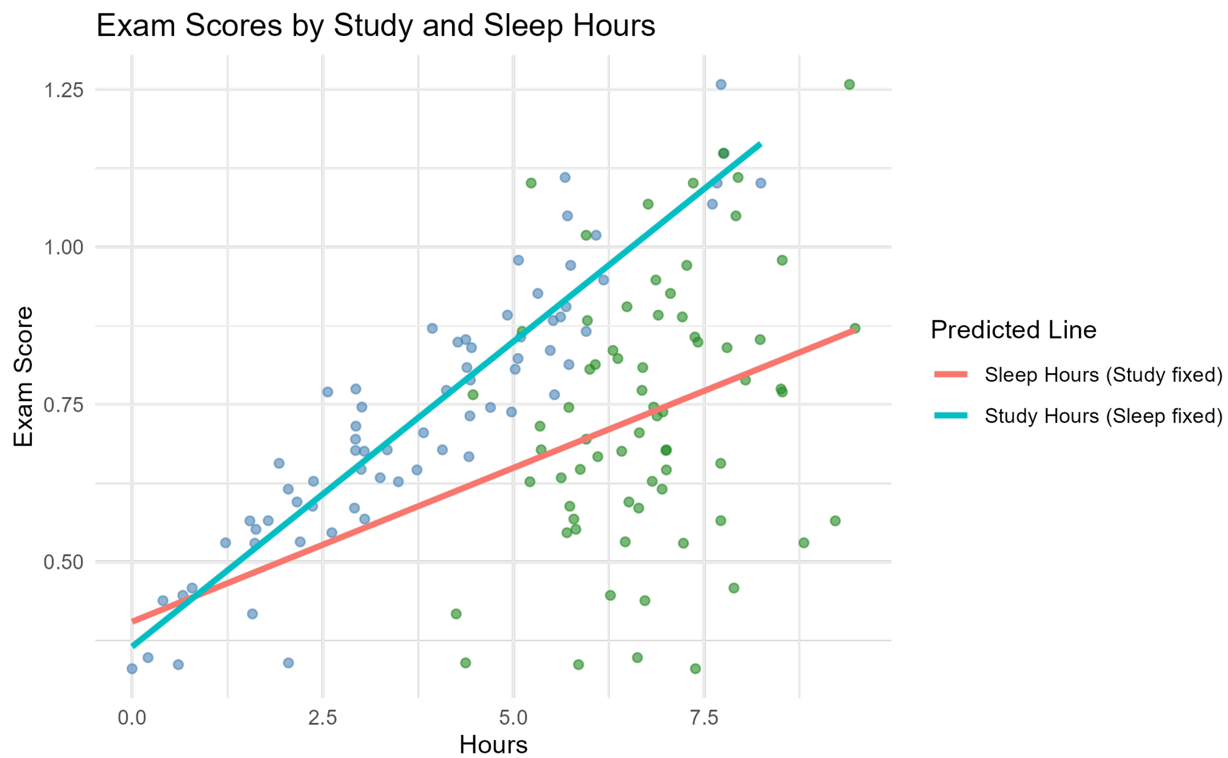
Something I do think is worthwhile is to show that, when you perform this regression analysis, we can show (graphically) the individual fitted lines we get. For example, here is our old friend exam scores with boht sleep and study hours!

2.4 Activity #1:

This one will be a bit more interactive, I just want to show you that the assumptions of the multiple regression model are fairly similar to that of the single.

Use the sheet I am giving out in class.

One point I want to hit just to cover Chapter 4: the R_j^2 reflects the relationship between the regressors themselves. A high degree of correlation leads to high R_j^2; what would this do to the denominator, and therefore the standard error? What should we conclude about highly similar regressors? This is known as multicollinearity.



2.5 Content Section #2:

Time:

2.5.1 R-squared

The R^2 concept discussed in the simple linear regression model applies effectively the same way in our multiple regression model. You'll recall that we can measure the variation in the dependent variable Y using the "total sum of squares." This can be decomposed into the residual and explained sum of squares:

$$TSS = \sum_i (Y_i - \bar{Y})^2 = \sum_i (\hat{Y}_i - \bar{Y})^2 + \sum_i \hat{u}_i{}^2 = ESS + RSS$$

and then we defined something called the R^2 which is the proportion of total variation that we can explain from our model:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}$$

One thing that I found pretty revolutionary was that the R^2 can *never* decrease when we add a variable. Intuitively, an additional variable can never *unexplain* the dependent variable. It can only not explain it at all, leading to zero effect.

I will leave it as an exericse to you to show exactly this: suppose you have two regressions, one with one independent variable and one with two. Suppose that the second variable in the second model has a coefficient of zero. Calculate the R^2 in each, and show that they are exactly the same.

In practice, if we have a regressor that explains none of the variation in the dependent variable, a natural question might be: why add it in at all? Take some time to consider the impact of adding a variable that has no explanatory power in light of what we know about efficiency and standard errors: what happens to the SE if we add a variable?

Inherent here is a tradeoff between additional variables and power. As we add more variables, unless those variables explain a good bit of the variation in our dependent variable, they may just make our regression noisy.

This motivates the next concept being the F test for inclusion of variables.

There are two F tests I will cover here:

1. Test of inclusion of all variables.
2. Test of including marginal variables.

In both we consider a general $k-$regressor equation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \tag{5}$$

In the **first** test, we are testing the following hypothesis:

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$

In words, this is testing if *none* of these variables should be included in the regression. This is very similar to the single regressor case where you were just testing whether or not β_1 actually has any explanatory power.

The F stat used here is:

$$F(k-1, n-1) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{6}$$

You will perform this test in the same way as before: there is a numerator and denominator degree of freedom to find the critical value (shown in parentheses), compare, then reject or fail to reject the null.

The **second** test considers a model in which we first regress the model above denoted by equation (5). We then consider whether we should add some more variables, specifically $m-k$ more variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \beta_{k+1} X_{k+1,i} + \cdots + \beta_m X_{mi} + u_i \tag{7}$$

The question here is, do these added (marginal) variables add anything of significance to our model? In hypothesis notation:

$$H_0 : \beta_{k+1} = \cdots = \beta_m = 0$$

where we use an F statistic of:

$$F(m-k, n-m) = \frac{(RSS_k - RSS_m)/(m-k)}{RSS_m/(n-m)} \tag{8}$$

where the RSS_k is the RSS from our first regression and RSS_m is the RSS from the second.

2.5.2 Quick Notes on "Prediction"

You will often associate the word "prediction" with something like forecasting, where we are predicting something into the future. That is a common practice in time series, and I encourage you to speak with Dr. Jeremy Piger if you have interest in that area of study.

In the case of cross-sectional data, where all we really care about is one time period, prediction is simply using the estimated coefficients to determine the outcome of a variable given a *new* set of observations.

For example, suppose that we are doing our estimation of sleep hours and study hours on exam scores and we get the following regression output:

Table 1:

	Dependent variable:
	exam_score
study_hours	0.097***
	(0.003)
sleep_hours	0.049***
	(0.006)
Constant	0.034
	(0.041)
Observations	70
R^2	0.935
Adjusted R^2	0.933
Residual Std. Error	0.054 (df = 67)
F Statistic	479.746*** (df = 2; 67)
Note:	*p<0.1; **p<0.05; ***p<0.01

Inputting our estimated coefficients, we yield the following model:

$$\text{exam\_score} = 0.034 + 0.097 * \text{study\_hours} + 0.049 * \text{sleep\_hours} \tag{9}$$

Now, suppose we had a student who was not part of this data set but had the same experience you all had, with me droning on about how cool econometrics is[1], and then took the test with study and sleep hours observed. We could "predict" the study and sleep hours of this individual based on what their study and sleep hours observed were.

Suppose a student named Connor had 4 hours of studying and a full 8 hours of sleep prior to taking the exam. We would expect them to get a score of:

$$0.034 + 0.097 * 4 + 0.049 * 8 = 0.814$$

[1] It really is!

2.6 Activity #2 (Optional):

Time:

2.7 Summary Activity (5 Q's of Reflection):

1.

3 Be Able To

- Write out a be able to - separate but related to the goals of this lecture.

4 Key Words

- Fill in Here