Multiple linear regression for ceteris paribus interpretation, functional form

Empirical Methods, Lecture 3

Tunga Kantarcı, FEB, Groningen University, Spring 2025

OLS, the ceteris paribus interpretation

OLS estimates of $\boldsymbol{\beta}_1$ in

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

and in

$$\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\upsilon}$$

are the same and given by

$$\hat{\boldsymbol{\beta}}_1^{OLS} = (\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)'}_{\boldsymbol{X}_1^{*\prime}}\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)}_{\boldsymbol{X}_1^{*}})^{-1}\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)'}_{\boldsymbol{X}_1^{*\prime}}\boldsymbol{y}.$$

Skip.

OLS, the ceteris paribus interpretation

$\boldsymbol{M}_2\boldsymbol{X}_1$ are the residuals from the regression of $\boldsymbol{X}_1$ on $\boldsymbol{X}_2$. To see this, note that

$$\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{P}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$$

where $\boldsymbol{P}_2$ is the projection matrix for $\boldsymbol{X}_2$. Post multiply by $\boldsymbol{X}_1$ to obtain

$$\boldsymbol{M}_2\boldsymbol{X}_1 = \underbrace{\boldsymbol{X}_1 - \underbrace{\boldsymbol{X}_2 \underbrace{(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1}_{\hat{\boldsymbol{\beta}}_{2,auxiliary}^{OLS}}}_{\hat{\boldsymbol{X}}_1}}_{\hat{\boldsymbol{\varepsilon}}}$$

where $(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_1$ are the OLS estimates on $\boldsymbol{X}_2$ in the regression of $\boldsymbol{X}_1$ on $\boldsymbol{X}_2$. This means that $\boldsymbol{M}_2$ projects $\boldsymbol{X}_1$ into the vector space that is orthogonal to the vector space spanned by $\boldsymbol{X}_2$. Hence, $\boldsymbol{M}_2\boldsymbol{X}_1 \perp \boldsymbol{X}_2$.

Skip.

OLS, the ceteris paribus interpretation

OLS estimates of $\boldsymbol{\beta}_1$ in

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

and in

$$\boldsymbol{y} = \boldsymbol{M}_2\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\upsilon}$$

are the same and given by

$$\hat{\boldsymbol{\beta}}_1^{OLS} = (\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)'}_{\boldsymbol{X}_1^{*\prime}}\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)}_{\boldsymbol{X}_1^{*}})^{-1}\underbrace{(\boldsymbol{M}_2\boldsymbol{X}_1)'}_{\boldsymbol{X}_1^{*\prime}}\boldsymbol{y}.$$

In the first model $\hat{\boldsymbol{\beta}}_1^{OLS}$ gives the effect of $\boldsymbol{X}_1$ on $\boldsymbol{y}$ controlling for the effect of $\boldsymbol{X}_2$. That is, $\boldsymbol{M}_2$ enters the formula of $\hat{\boldsymbol{\beta}}_1^{OLS}$! This is the power of the multiple regression analysis. It allows to do in a nonexperimental economic setting what natural scientists are able to do in a controlled laboratory setting: keeping other factors fixed. It provides this ceteris paribus interpretation although the data have not been collected in a ceteris paribus fashion. Skip.

OLS, the ceteris paribus interpretation

Let's study this in level form.

Consider the liner model with one explanatory variable:

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

We have shown that the OLS estimator of β_1 is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

OLS, the ceteris paribus interpretation

Consider a new model where a second explanatory variable is added:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i.$$

In this case the OLS estimator of β_1 is given by

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

where $\hat{r}_{i1}$ are the OLS residuals from a regression of x_{i1} on x_{i2}. That is, $\hat{r}_{i1}$ represents transformed x_{i1} where the impact of x_{i2} on x_{i1} is netted out from x_{i1}. We then regress y_i on $\hat{r}_{i1}$ which gives the impact of x_{i1} on y_i where x_{i2} plays no role. Hence, in multiple regression analysis, β_1 gives the effect of x_{i1} on y_i controlling for the effect of other variables, x_{i2} in this case.

OLS, the ceteris paribus interpretation

This is the power of the multiple regression analysis. It allows to do in a non-experimental economic setting what natural scientists are able to do in a controlled laboratory setting: keeping other factors fixed. It provides this ceteris paribus interpretation although the data have not been collected in a ceteris paribus fashion.

This result is fundamental in regression analysis and originates from publications in the first issue of Econometrica based on Frisch and Waugh (1933) and Lovell (1963) which led to the FWL theorem.

Yet, this does not guarantee a causal effect. It guarantees conditional correlation.

OLS, the ceteris paribus interpretation, example

Consider the regression of ***wage*** on ***educ***

```
. regress wage educ

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    251.46
       Model |  7842.35455         1  7842.35455   Prob > F        =    0.0000
    Residual |  31031.0745       995  31.1870095   R-squared       =    0.2017
-------------+----------------------------------   Adj R-squared   =    0.2009
       Total |   38873.429       996  39.0295472   Root MSE        =    5.5845

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   1.135645   .0716154    15.86   0.000     .9951106     1.27618
       _cons |  -4.860424   .9679821    -5.02   0.000    -6.759944   -2.960903
------------------------------------------------------------------------------
```

OLS, the ceteris paribus interpretation, example

Consider the regression of ***wage*** on ***educ*** and ***exper***

```
. regress wage educ exper
```

Source	SS	df	MS
Model	10008.3629	2	5004.18147
Residual	28865.0661	994	29.0393019
Total	38873.429	996	39.0295472

Number of obs	=	997
F(2, 994)	=	172.32
Prob > F	=	0.0000
R-squared	=	0.2575
Adj R-squared	=	0.2560
Root MSE	=	5.3888

wage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	1.246932	.0702966	17.74	0.000	1.108985	1.384879
exper	.1327808	.0153744	8.64	0.000	.1026108	.1629509
_cons	-8.833768	1.041212	-8.48	0.000	-10.87699	-6.790542

The coefficient of ***educ*** has changed, signalling that ***educ*** and ***exper*** are correlated, and that we should control for ***exper*** in our model.

OLS, the ceteris paribus interpretation, example

Consider the regression of ***educ*** on ***exper***

```
. regress educ exper

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =     34.59
       Model |  204.317954         1  204.317954   Prob > F        =    0.0000
    Residual |  5876.48847       995  5.90601856   R-squared       =    0.0336
-------------+----------------------------------   Adj R-squared   =    0.0326
       Total |  6080.80642       996  6.10522733   Root MSE        =    2.4302

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       exper |  -.0400901    .006816    -5.88   0.000    -.0534655   -.0267147
       _cons |   14.04201   .1493993    93.99   0.000     13.74884    14.33519
------------------------------------------------------------------------------
```

educ and ***exper*** are negatively correlated, which explains why the coefficient of ***educ*** increased when we controlled for ***exper***. Obtain the residuals of this model, and call them ***Reduc***:

```
. predict Reduc, resid
```

OLS, the ceteris paribus interpretation, example

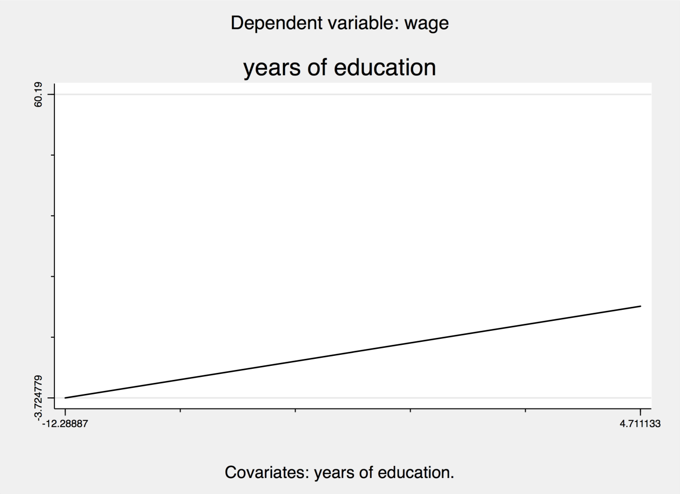
Consider the regression of ***wage*** on ***Reduc***

```
. regress wage Reduc

      Source |       SS           df       MS      Number of obs   =       997
-------------+----------------------------------   F(1, 995)       =    305.73
       Model |  9136.99562         1  9136.99562   Prob > F        =    0.0000
    Residual |  29736.4334       995  29.8858627   R-squared       =    0.2350
-------------+----------------------------------   Adj R-squared   =    0.2343
       Total |   38873.429       996  39.0295472   Root MSE        =    5.4668

------------------------------------------------------------------------------
        wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       Reduc |   1.246932   .0713139    17.49   0.000     1.106989    1.386875
       _cons |   10.23101   .1731352    59.09   0.000     9.891261    10.57077
------------------------------------------------------------------------------
```

The coefficient of ***Reduc*** in this regression and the coefficient of ***educ*** in the full model considered above are the same, as the FWL theorem requires.

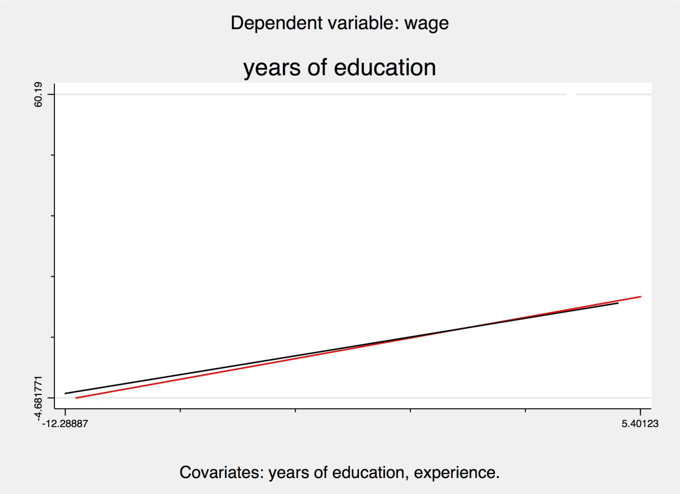
OLS, the ceteris paribus interpretation, example

The figure shows the fitted line from the regression of ***wage*** on ***educ***.



OLS, the ceteris paribus interpretation, example

Adding to the figure the fitted line from the regression of ***wage*** on ***educ*** after partialling out the effect of ***exper*** (red line).



SLM, functional form, quadratic in the variable

A regression model that is quadratic in a regressor is often of interest because it provides an economic interpretation. It provides the decreasing marginal utility interpretation.

SLM, functional form, quadratic in the variable

Consider the regression model

$$wage_i = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + u_i$$

which is quadratic in age.

Assume that

$$\frac{\partial u_i}{\partial age_i} = 0$$

such that u_i contains no information about age_i.

SLM, functional form, quadratic in the variable

If age changes by a small amount, the dependent changes approximately by the amount

$$\frac{\partial wage_i}{\partial age_i} = \beta_1 + 2\beta_2 age_i.$$

If age changes by some discrete amount, the dependent changes exactly by

$$\begin{aligned}\Delta wage_i &= (\beta_0 + \beta_1(age_i + \Delta age_i) + \beta_2(age_i + \Delta age_i)^2) \\ &\quad - (\beta_0 + \beta_1 age_i + \beta_2 age_i^2).\end{aligned}$$

The approximate change will be good for small changes but bad for big changes in *age*.

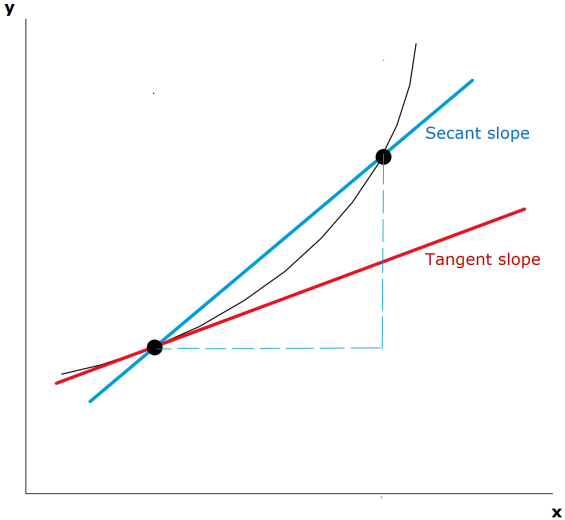
SLM, functional form, quadratic in the variable

In reference to calculus, the former change is the tangent slope, and the latter is the secant slope. The tangent slope refers to a marginal change, and the secant slope refers to a discrete change in the nonlinear function. We take the tangent slope as an approximation to the secant slope. That is,

$$\frac{\delta y}{\delta x} \approx \frac{\Delta y}{\Delta x}.$$

SLM, functional form, quadratic in the variable

The smaller is the secant slope, the closer it is to the tangent slope, then the tangent slope approximates the secant slope better:



SLM, functional form, quadratic in the variable

In either case, the expression states that the partial effect of *age*, or the slope of the relationship between *wage* and *age*, depends on the value of *age*.

SLM, functional form, quadratic in the variable

By taking the derivative of the *wage* function with respect to *age* and setting it equal to zero, we can find the critical point where the *wage* function could have a maximum or minimum. That is,

$$\frac{\partial wage_i}{\partial age_i} = \beta_1 + 2\beta_2 age_i = 0$$

and solving for *age*,

$$age_{i,max} = \frac{-\beta_1}{2\beta_2}.$$

So, the *age* at which the *wage* reaches its maximum (or minimum) is determined by this critical point. If β_2 is negative, this point will be a maximum. If it is positive, it will be a minimum.

SLM, functional form, logarithmic in the variable

The logarithmic function also provides an economic interpretation.

It also has econometric implications but we do not cover them here. For example, it has implications if there is heteroskedasticity.

SLM, functional form, logarithmic in the variable

Consider the following logarithmic change:

$$\ln(41) - \ln(40) \approx 0.024.$$

The proportionate change, or relative change, we frequently calculate is

$$\frac{41 - 40}{40} = 0.025.$$

The two quantities are very close. This shows that, for small changes, the logarithmic change closely approximates the proportionate change.

SLM, functional form, logarithmic in the variable

Consider the following logarithmic change:

$$\ln(60) - \ln(40) \approx 0.405.$$

The proportionate change is

$$\frac{60 - 40}{40} = 0.500.$$

The two quantities are not really close. This shows that, for large changes, the approximation is not accurate.

SLM, functional form, logarithmic in the variable

We can make use of this result in regression analysis. Consider the log-linear model

$$\ln(y_i) = \beta_1 x_{i1} + u_i.$$

We are interested in the change in $\ln(y_i)$ when we change x_{i1} by some unit.

Assume that

$$\frac{\partial u_i}{\partial x_{i1}} = 0$$

such that u_i contains no information about x_{i1}.

SLM, functional form, logarithmic in the variable

$$\ln(y_i) = \beta_1 x_{i1} + u_i.$$

For small changes in x_{i1}, the change in $\ln(y_i)$ closely approximates the proportionate change in y. In this case we can consider the derivate

$$\frac{\partial \ln(y_i)}{\partial x_{i1}} = \beta_1.$$

This change refers to the tangent slope.

SLM, functional form, logarithmic in the variable

$$\ln(y_i) = \beta_1 x_{i1} + u.$$

For large changes in x_1, the approximation is worse. In this case the exact proportionate change can be calculated as

$$\Delta \ln(y_i) = \ln(y_{i1}) - \ln(y_{i0}) = \beta_1 \Delta x_{i1}.$$

Taking the terms to the exponential gives

$$\begin{aligned}
e^{(\ln(y_{i1}) - \ln(y_{i0}))} &= e^{\beta_1 \Delta x_{i1}} \\
e^{\ln \frac{y_{i1}}{y_{i0}}} &= e^{\beta_1 \Delta x_{i1}} \\
\frac{y_{i1}}{y_{i0}} &= e^{\beta_1 \Delta x_{i1}} \\
\frac{y_{i1}}{y_{i0}} - 1 &= e^{\beta_1 \Delta x_{i1}} - 1 \\
\frac{y_{i1} - y_{i0}}{y_{i0}} &= e^{\beta_1 \Delta x_{i1}} - 1
\end{aligned}$$

This change refers to the secant slope.

SLM, functional form, logarithmic in the variable

The main implications of the logarithmic transformation for applied work are the following. In the log-linear model

$$\ln(y_i) = \beta_1 x_{i1} + u_i,$$

we have

$$\frac{\Delta \ln(y_i)}{\Delta x_{i1}} = \beta_1$$

which gives a proportionate change interpretation. If we multiply by 100, we have

$$100 * \frac{\Delta \ln(y_i)}{\Delta x_{i1}} = 100 * \beta_1$$

which gives a percentage change interpretation. That is,

$$\frac{\% \Delta y_i}{\Delta x_{i1}} = 100 * \beta_1.$$

SLM, functional form, logarithmic in the variable

In the linear-log model

$$y_i = \beta_1 \ln\left(x_{i1}\right) + u_i$$

we have

$$\frac{1}{100} * \frac{\Delta y_i}{\Delta \ln\left(x_{i1}\right)} = \frac{1}{100} * \beta_1$$

and

$$\frac{\Delta y_i}{\% \Delta x_{i1}} = \frac{\beta_1}{100}$$

SLM, functional form, logarithmic in the variable

In the log-log model

$$\ln(y_i) = \beta_1 \ln(x_{i1}) + u,$$

we have

$$\frac{100}{100} * \frac{\Delta \ln(y_i)}{\Delta \ln(x_{i1})} = \beta_1$$

and

$$\frac{\Delta \ln(y_i)}{\Delta \ln(x_{i1})} = \beta_1.$$

This is constant elasticity. It is often used in applied work. In all cases, the proportionate change is approximated by the logarithmic change.

SLM, functional form, interaction effect

Suppose that the model of interest is

$$wage_i = \beta_0 + \beta_1 educ_i + \beta_2 female_i + \beta_3 educ_i * female_i + u_i.$$

$educ_i$ is a continuous variable.

$female_i$ is a dummy variable.

$educ_i * female_i$ is an interaction variable which indicates $educ_i$ with respect to the groups of $female_i$.

SLM, functional form, interaction effect

We can see the motivation behind an interaction variable if we change either of the two interacting variables.

SLM, functional form, interaction effect

First, consider a unit change in *educ*:

$$\frac{\Delta \mathit{wage}_i}{\Delta \mathit{educ}_i} = \beta_1 + \beta_3 \mathit{female}_i$$

This equation indicates that the effect of *educ* on *wage* depends on the group *female*.

SLM, functional form, interaction effect

Second, consider a change in *female*.

If $female = 1$, we obtain the following model:

$$wage_i = \beta_0 + \beta_2 + (\beta_1 + \beta_3)educ_i.$$

If $female = 0$, we obtain another model:

$$wage_i = \beta_0 + \beta_1 educ_i.$$

The two models differ in the constant β_2 and the slope β_3. We get the same interpretation: The effect of *educ* on *wage* depends on the *female*.

SLM, functional form, interaction effect

$educ * female$ is also called the *slope dummy variable* because it allows a change in the slope.