Econometrics 1
Lecture 3: Multiple Linear Regression
ENSAE 2014/2015

Michael Visser (CREST-ENSAE)

Definition of MLR

Suppose we wish to investigate the relation between

- An outcome variable y (the regressand, dependent variable, etc.)
- A set of k variables x_1, x_2, ..., x_k (the regressors, covariates, etc.)

The Multiple Linear Regression (MLR) model assumes that the following relationship holds in the population

Linear in parameters

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

Remarks:

- As in the SLR model, β_0 is the constant (or intercept), and β_1, ..., β_k are slope parameters.
- The error term u captures the combined effect of other variables not included in the model.
- The model is linear in the parameters (not necessarily in the explanatory variables).

Definition of MLR (cntd)

- y, x_1, ..., x_k, u are viewed as random variables.
- The model reflects a *population* relationship. A sample from the population will be used to learn (i.e. estimate) something about the parameter values in the population.
- β_0 is the value of y when all x's and u equal zero. β_k is the causal effect of a unit-change in x_k on y, i.e., it measures the marginal effect of x_k on y keeping u and x_j, $j \neq k$, fixed.

Advantages of MLR model (compared to SLR):

- MLR models are obviously richer. For instance the following model (*exper* is years of experience on labor market, *tenure* measures years of experience with current employer)

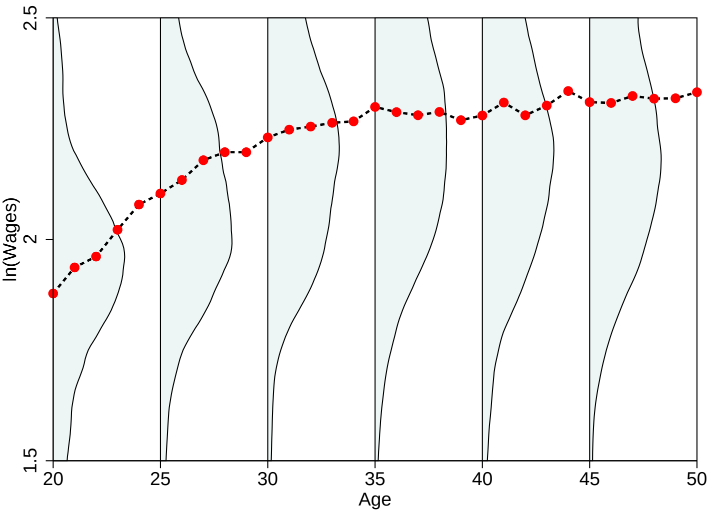
$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u \tag{1}$$

  allows the researcher to investigate not just the effect of education (as in the SLR model $wage = \beta_0 + \beta_1 educ + u$), but also of experience and tenure.

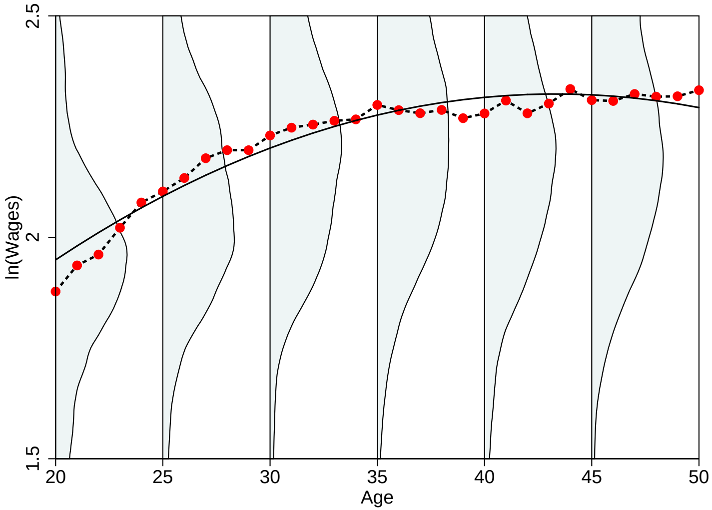
Definition of MLR (cntd)

- Adding additional explanatory variables leads to a better fit, which is important if one wishes to use the estimations for predictions.
- Adding additional regressors augments the likelihood that the Zero Conditional Mean assumption holds. This increases in turn the likelihood that we identify causal effects. Example: in the SLR model $wage = \beta_0 + \beta_1 educ + u$, we need to assume that experience is not correlated with education. In the MLR model (1) experience (and also tenure) are taken out of the error term, so this assumption is no longer necessary.
- MLR models allow for more general specifications. For instance a specific variable may have a non-linear effect on y : $log(wage) = \beta_0 + \beta_1 age + \beta_2 age^2 + u$. (note that $\beta_1$does not measure the causal effect of age since age^2 cannot be kept fixed when age is changed. Instead $\triangle log(wage)/\triangle age = \beta_1 + 2\beta_2 age$.) See next figure.

Age profile wages



Age profile wages



Assumptions

- MLR.1 Linear In Parameters: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$
- MLR.2 Random Sample: $\{(x_{i1}, \ldots, x_{ik}, y_i) : i = 1, \ldots, n\}$, where $\{x_{i1}, \ldots, x_{ik}, y_i\}$ are *i.i.d.*
- MLR.3 Zero Conditional Mean: $E[u|x_1, \ldots, x_k] = 0$
- MLR.4 No perfect collinearity: $x_j \neq c$ and no exact linear relationships among the x_j in the sample
- MLR.5 Homoscedasticity: $Var[u|x_1, \ldots, x_k] = E[u^2|x_1, \ldots, x_k] = \sigma^2$

Remarks:

- Straightforward extensions of assumption in SLR model.
- MLR.3 implies that each variable x_j is uncorrelated with u : $E[u|x_j] = E[E[u|x_1, \ldots, x_k]] = 0$ (first expectation is wrt to all $x's$ given x_j).

Assumptions (cntd)

- If MLR.3 holds we say that the x's are exogenous. Otherwise they are said to be endogenous.
- MLR.4 not only says that each x_j should vary in the sample, but also that they should not be perfectly linearly related. Example: in the model (*man* equals 1 if the individual is a man, and 0 otherwise; *woman* is defined in similar way)

$$wage = \beta_0 + \beta_1 man + \beta_2 woman + u$$

there is a linear relationship between the variables *man* and *woman* ($man + woman = 1$).

Estimation

As in the SLR model, we can obtain the estimator of $\beta_0, \beta_2, ..., \beta_k$ in two ways. In the first method we use the $k+1$ moment conditions (that follow from MLR.3)

$$E[u] = E[y - \beta_0 - \beta_1 x_1 - \ldots - \beta_k x_k] = 0$$
$$E[x_j u] = E[x_j(y - \beta_0 - \beta_1 x_1 - \ldots - \beta_k x_k)] = 0,\ j = 1, 2, \ldots, k$$

and the estimator $\hat{\beta}_0, \ldots, \hat{\beta}_k$ is obtained by imposing that the corresponding empirical sample means equal zero.
In the second method the estimator is obtained by minimizing the sum of squared residuals:

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_k x_{ik})^2$$

with respect to β. Both approaches lead to exactly the same estimator, and is called the Ordinary Least Squares (OLS) estimator of $\beta_0, \beta_2, ..., \beta_k$.

OLS 1st order conditions

$$
\begin{aligned}
\frac{\partial Q}{\partial \beta_0} &= -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0 \\
\frac{\partial Q}{\partial \beta_1} &= -2\sum_{i=1}^{n}x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0 \\
&\vdots \\
\frac{\partial Q}{\partial \beta_k} &= -2\sum_{i=1}^{n}x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0
\end{aligned}
$$

Under MLR.4 this system of $k+1$ equations and $k+1$ unknowns has a unique solution.

Definition of OLS estimator

The solution of the above system of first-order conditions is the OLS estimator.

OLS Estimator

$$\begin{aligned}\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1 \overline{x}_1 - \ldots - \hat{\beta}_k \overline{x}_k \\ \hat{\beta}_j &= \frac{\sum_{i=1}^{n} \hat{r}_{ij} y_i}{\sum_{i=1}^{n} \hat{r}_{ij}^2}, \qquad j = 1, 2, \ldots, k\end{aligned}$$

where $\bar{x}_j = n^{-1} \sum_{i=1}^{n} x_{ij}$ and $\hat{r}_{ij}$ the OLS residuals from a regression of x_j on the other explanatory variables and a constant.

Definition of OLS estimator (cntd)

We can rewrite $\hat{\beta}_j$ as (since the average of residuals equals zero: $\bar{\hat{r}}_j = \frac{1}{n}\sum_i \hat{r}_{ij} = 0$):

$$\hat{\beta}_j = \frac{\sum_{i=1}^n \hat{r}_{ij} y_i}{\sum_{i=1}^n \hat{r}_{ij}^2} = \frac{\sum_{i=1}^n (\hat{r}_{ij} - \bar{\hat{r}}_j)(y_i - \bar{y})}{\sum_{i=1}^n (\hat{r}_{ij} - \bar{\hat{r}}_j)^2}, \, j = 1, 2, \ldots, k$$

So, recalling the definition of the OLS estimator of the slope parameter in the SLR model, we see that $\hat{\beta}_j$ is in fact the estimator of the slope parameter in a regression of y on a constant and $\hat{r}_j$. The estimate $\hat{\beta}_j$ can thus be seen as the effect of x_j on y after the other x's are partialled out ($\hat{r}_{ij}$ can be interpreted as the "part" of x_{ij} that is left after we take out the variation in x_{il}, $l \neq j$.)

OLS - Two explanatory variables

To show that the OLS estimator is unbiased it is helpful to consider first the model with just two explanatory variables ($k = 2$). We thus want to estimate the following MLR

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

We will check that $\hat{\beta}_1$ is unbiased (the proof is similar for $\hat{\beta}_2$). First estimate the following SLR

$$x_{i1} = \alpha_0 + \alpha_1 x_{i2} + r_{i1}$$

and use the OLS estimates $\hat{\alpha}_0, \hat{\alpha}_1$ to construct the residual:

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

which is the "part" of x_{i1} not explained by x_{i2} .

OLS - Two explanatory variables (cntd)

The residual $\hat{r}_{i1} = x_{i1} - \hat{\alpha}_0 - \hat{\alpha}_1 x_{i2}$ has the usual properties (this follows form the first-order conditions)

- sum to zero: $\sum_{i=1}^{n} \hat{r}_{i1} = 0$
- orthogonal to regressors:$\sum_{i=1}^{n} \hat{r}_{i1} x_{i2} = 0$

And it also follows that

$$\begin{aligned}
\sum_{i=1}^{n} \hat{r}_{i1} x_{i1} &= \sum_{i=1}^{n} \hat{r}_{i1}(x_{i1} - \hat{\alpha}_0 - \hat{\alpha}_1 x_{i2} + \hat{\alpha}_0 + \hat{\alpha}_1 x_{i2}) \\
&= \sum_{i=1}^{n} \hat{r}_{i1}(\hat{r}_{i1} + \hat{\alpha}_0 + \hat{\alpha}_1 x_{i2}) \\
&= \sum_{i=1}^{n} \hat{r}_{i1}^2 + \hat{\alpha}_0 \sum_{i=1}^{n} \hat{r}_{i1} + \hat{\alpha}_1 \sum_{i=1}^{n} \hat{r}_{i1} x_{i2} = \sum_{i=1}^{n} \hat{r}_{i1}^2
\end{aligned}$$

OLS - Two explanatory variables (cntd)

The OLS estimator of β_1 is the OLS estimator of the slope parameter in the regression of y on a constant and $\hat{r}_{i1}$:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(\hat{r}_{i1} - \bar{\hat{r}}_1)y_i}{\sum_{i=1}^{n}(\hat{r}_{i1} - \bar{\hat{r}}_1)^2} = \frac{\sum_{i=1}^{n}\hat{r}_{i1}y_i}{\sum_{i=1}^{n}\hat{r}_{i1}^2} \\
&= \frac{\sum_{i=1}^{n}\hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i)}{\sum_{i=1}^{n}\hat{r}_{i1}^2} \\
&= \beta_1 + \frac{\sum_{i=1}^{n}\hat{r}_{i1}u_i}{\sum_{i=1}^{n}\hat{r}_{i1}^2}
\end{aligned}$$

which is an unbiased estimator of β_1: $E[\hat{\beta}_1|(x_{i1}, x_{i2}), \forall i] = \beta_1$, which implies $E[\hat{\beta}_1] = \beta_1$.

Unbiasedness

To show that the OLS estimator is unbiased in the general case we proceed in an analogous way:

$$\hat{\beta}_j = \frac{\sum_{i=1}^n \hat{r}_{ij} y_i}{\sum_{i=1}^n \hat{r}_{ij}^2} \overset{MLR.1}{=} \beta_j + \frac{\sum_{i=1}^n \hat{r}_{ij} u_i}{\sum_{i=1}^n \hat{r}_{ij}^2}$$

since $\sum_{i=1}^n \hat{r}_{ij} = 0$, $\sum_{i=1}^n x_{il}\hat{r}_{ij} = 0$, $\forall l \neq j$ and $\sum_{i=1}^n x_{ij}\hat{r}_{ij} = \sum_{i=1}^n \hat{r}_{ij}^2$. Now

$$\begin{aligned}
E[\hat{\beta}_j|(x_{1i},\ldots,x_{ki})\forall i] &= \beta_j + \frac{\sum_{i=1}^n \hat{r}_{ij} E[u_i|(x_{i1},\ldots,x_{ik})\forall i]}{\sum_{i=1}^n \hat{r}_{ij}^2} \\
&\overset{MLR.2}{=} \beta_j + \frac{\sum_{i=1}^n \hat{r}_{ij} E[u_i|x_{i1},\ldots,x_{ik}]}{\sum_{i=1}^n \hat{r}_{ij}^2} \\
&\overset{MLR.3}{=} \beta_j
\end{aligned}$$

Unbiasedness (cntd)

Unbiasedness of OLS

Under assumptions MLR.1 to MLR.4,

$$E[\hat{\beta}_j] = \beta_j, \quad j = 0, 1, \ldots, k$$

Remarks:

- The unbiasedness of $\hat{\beta}_0$ can be shown as in the SLR model (see slides of last week).
- The results are true for any values of the population parameters β_j. Therefore, including *irrelevant* variables ($\beta_l = 0$) does not affect the unbiasedness of the estimator. (It can affect however the variances of the OLS estimators.)

Unbiasedness (cntd)

- Omitting *relevant* regressors obviously <u>does</u> in general affect the estimators since this violates assumption MLR.3 if the omitted regressor is correlated with the variables included in the model.
- Unbiasedness does not mean that the OLS estimate obtained using a particular sample necessarily corresponds to the true population parameter. Although the estimate and true value may of course coincide, this will generally not be the case (see simulation study of last week).

Fitted values, residuals, and Goodness-of-fit

As in the SLR model we can calculate the fitted value

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

and the residual

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

As in the simple regression model the R^2 measures the fraction of the sample variation in y explained by the explanatory variables:

$$\begin{aligned} R^2 &= \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \\ &= \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n} \hat{u}_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \end{aligned}$$

Remarks:

- R^2 never decreases as we add more variables to the regression. This mechanical increase makes that the R2 *as such* is NOT a good way to decide whether we need to add variables to the model. We can use hypothesis testing (next week) to determine this.

Fitted values, residuals, and Goodness-of-fit (cntd)

- Low R2 does not say anything about whether we estimate causal/ceteris paribus effects.
- Low R2 does say that the model is not so useful for prediction.

Omitted variable bias

Suppose true model is

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

but that for some reason (for example because x_2 is not recorded in the data set) we regress y only on x_1 and a constant (i.e., x_2 is not included in the model). Letting $\widetilde{\beta}_0$ and $\widetilde{\beta}_1$ be the OLS estimators, and $\tilde{y}$ the fitted value of y, the OLS regression line is

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

and

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

we can now replace y_i with the true model and get

$$\tilde{\beta}_1 = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)^2}$$

Omitted variable bias (contd)

Taking expectations of $\tilde{\beta}_1$

$$E[\tilde{\beta}_1 | x_{1i}, x_{2i}, \forall i] = \beta_1 + \beta_2 \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

and we see that we get a biased estimate of β_1. The above fraction (denoted $\tilde{\delta}_1$) is the slope estimate from a regression of x_2 on a constant and x_1

$$\text{Bias} = E[\tilde{\beta}_1 - \beta_1 | x_{1i}, x_{2i}, \forall i] = \beta_2 \tilde{\delta}_1$$

Bias depends on β_2 and correlation between x_1 and x_2:

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

Omitted variable bias (contd)

Remarks:

- **Sign** of bias depends both on sign β_2 and sign $corr(x_1, x_2)$.
- If Bias>0 we say that we have an *upward* bias, if Bias<0 a *downward* bias.
- **Size** of bias depends on size of β_2 and $\tilde{\delta}_1$.
- Although they are unknown, we usually have priors on *signs* of β_2 and $\tilde{\delta}_1$. Size is more problematic.
- There is no bias only if $\beta_2 = 0$ (x_2 not a relevant variable) and/or $\tilde{\delta}_1 = 0$ (x_1 and x_2 uncorrelated).

Omitted variable bias: k>2

- In the general case ($k > 2$), deriving signs on the bias is more complicated because it involves all pairwise correlations and how these correlations are weighted with the β's.
- Correlation between a *single* regressor and the error term usually results in *all* OLS estimators being biased.

Example: Omitted Variable Bias

Last week we presented an estimation of the attendance/exam model (ENSAE data):

$$exam = \beta_0 + \beta_1 presence + u$$

The question we wish to address: is the estimate of β_1 biased by omitted variables ?

One possible missing variable: scholastic aptitude. It is likely that well-performing students get high exam scores (whether they go to exercise classes or not). More able students may also go to exercise class more often.

The data set records the average grade obtained in the first year of ENSAE. This is arguably a good proxy for scholastic aptitude.

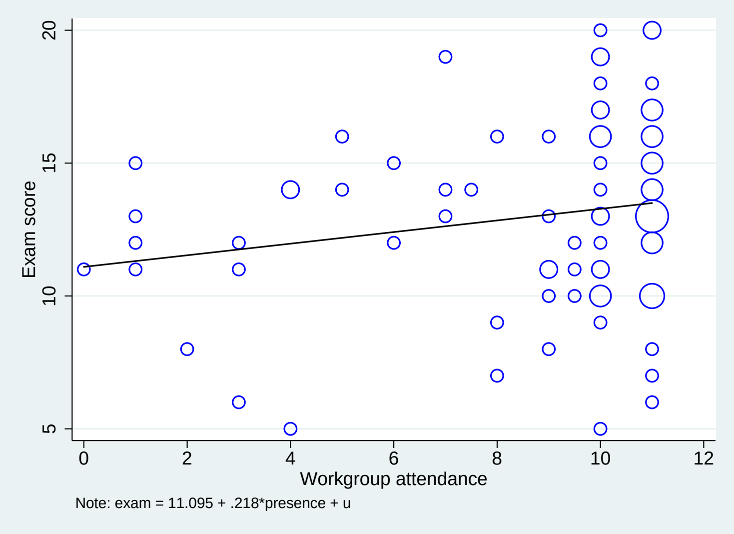
Problem: this variable ($moy1a$) is not observed for all students (e.g., those who entered ENSAE directly as 2nd-year students).

The sample size is therefore reduced from $n = 119$ to $n = 85$.

The following empirical analysis shows that there is indeed evidence that last week's results are biased.

Example: Omitted Variable Bias (cntd)

Attendance & performance - students who did 1st year ENSAE



Example: Omitted Variable Bias (cntd)

Regression performance on attendance - students who did 1st year ENSAE

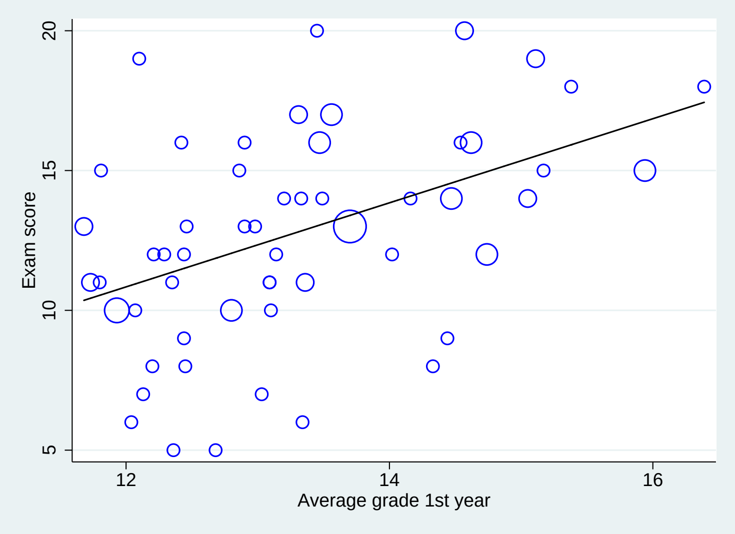
```
. reg exam presence

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  1,    83) =    3.13
       Model |  37.2835482     1  37.2835482           Prob > F      =  0.0804
    Residual |  987.704687    83  11.9000565           R-squared     =  0.0364
-------------+------------------------------           Adj R-squared =  0.0248
       Total |  1024.98824    84  12.2022409           Root MSE      =  3.4496

------------------------------------------------------------------------------
        exam |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    presence |   .2183855   .1233786     1.77   0.080    -.0270095    .4637806
       _cons |   11.09511   1.145652     9.68   0.000     8.816455    13.37377
------------------------------------------------------------------------------
```

Example: Omitted Variable Bias (cntd)

Performance and 1st year grade



Example: Omitted Variable Bias

Regression performance on 1st year grades

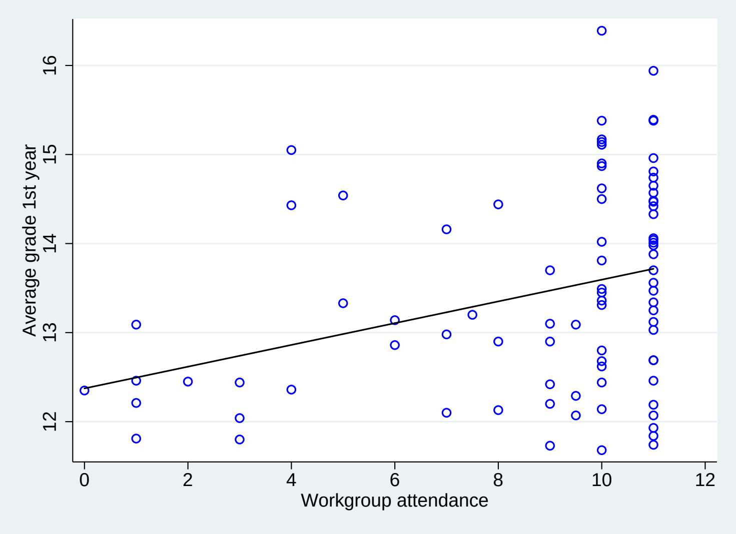
```
. reg exam moy

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  1,    83) =   27.11
       Model |  252.390525     1  252.390525           Prob > F      =  0.0000
    Residual |   772.59771    83  9.30840615           R-squared     =  0.2462
-------------+------------------------------           Adj R-squared =  0.2372
       Total |  1024.98824    84  12.2022409           Root MSE      =   3.051

------------------------------------------------------------------------------
        exam |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       moy1a |   1.503445   .2887278     5.21   0.000     .9291768    2.077713
       _cons |  -7.202315   3.896076    -1.85   0.068    -14.95145    .5468225
------------------------------------------------------------------------------
```

Example: Omitted Variable Bias (cntd)

1st year grades and attendance



Example: Omitted Variable Bias (cntd)

Regression 1st year grades on attendance

```
. reg moy presence

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  1,    83) =    9.68
       Model |  11.6598981     1  11.6598981           Prob > F      =  0.0026
    Residual |  100.000222    83  1.20482195           R-squared     =  0.1044
-------------+------------------------------           Adj R-squared =  0.0936
       Total |   111.66012    84  1.32928714           Root MSE      =  1.0976

------------------------------------------------------------------------------
       moy1a |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    presence |   .1221273   .0392579     3.11   0.003     .0440449    .2002096
       _cons |   12.37333   .3645353    33.94   0.000     11.64828    13.09838
------------------------------------------------------------------------------
```

Example: Omitted Variable Bias (cntd)

What do we expect?

Remember

$$\text{Bias} = \beta_2 \tilde{\delta}_1$$

One expects the "effect" of 1st year grades on the exam score to be positive (one of the above above regressions shows that 1st year grade has a positive effect on exam score):

$$\beta_2 > 0$$

We have also seen that 1st year grades and exercise class attendance are positively correlated:

$$\tilde{\delta}_1 \approx 0.122 > 0$$

The bias should therefore be positive and our initial estimate (0.218) too high ...

Example: Omitted Variable Bias

Attendance regression, including 1st year grades

```
. reg exam presence moy

      Source |       SS       df       MS              Number of obs =      85
-------------+------------------------------           F(  2,    82) =   13.47
       Model |  253.446064     2  126.723032           Prob > F      =  0.0000
    Residual |  771.542171    82  9.40905087           R-squared     =  0.2473
-------------+------------------------------           Adj R-squared =  0.2289
       Total |  1024.98824    84  12.2022409           Root MSE      =  3.0674

------------------------------------------------------------------------------
        exam |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    presence |   .0388285   .1159276     0.33   0.739    -.1917883    .2694454
       moy1a |   1.470245   .3067414     4.79   0.000     .8600386    2.080451
       _cons |  -7.096716    3.92975    -1.81   0.075    -14.91424    .7208074
------------------------------------------------------------------------------
```

Variance of OLS estimators

Using assumptions MLR.1 to MLR.4 and the homoscedasticity assumption MLR.5 we can derive the following result:

Sampling Variance of OLS estimators

$$Var(\hat{\beta}_j|X) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$$

for $j = 1, 2, \ldots, k$, where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ and R_j^2 is the R-squared from a regression of x_j on the other regressors (and an intercept)

Variance of OLS estimators (cntd)

Remarks:

- The variance of the estimator increases with the error variance and decreases with the variance in the regressor.
- Since the total sample variation in x (SST) increases with sample size, the variance of the estimator decreases with the sample size.
- R_j^2 is the fraction in the sample variation in x_{ij} explained by other regressors. If another (possibly irrelevant) variable is added to the model, R_j^2 increases, and $Var(\hat{\beta}_j|X)$ increases.
- $R_j^2 = 0$ means that all slope coefficients (in the regression of x_j on a constant and all other regressors) equal zero: x_j is uncorrelated with every other regressor. The variance of $\hat{\beta}_j$ is then minimal (occurs very rarely in practice).

Variance of OLS estimators (cntd)

- $R_j^2 = 1$ means that there is a perfect linear relationship between x_j and (some of) the other regressors, a case that is excluded by MLR.4.
- In practice, $0 < R_j^2 < 1$. When it is close to 1, the variance can get very large. This reflects that x_j is highly correlated with (some of) the regressors, a problem that is called multicollinearity.
- As in SLR analysis, the variances of OLS estimators depend on σ which is generally unknown. But it can be estimated using the data.

Derivation: Variance OLS estimator

From above we have

$$\hat{\beta}_j = \beta_j + \frac{\sum_{i=1}^n \hat{r}_{ij} u_i}{\sum_{i=1}^n \hat{r}_{ij}^2}$$

And we derive the variance as follows

$$\begin{aligned}
Var(\hat{\beta}_j|(x_{1i}, \ldots, x_{ki})\forall i) &= Var((\sum_{i=1}^{n} \hat{r}_{ij}^2)^{-1} \sum_{i=1}^{n} \hat{r}_{ij} u_i|(x_{1i}, \ldots, x_{ki})\forall i) \\
&= (\sum_{i=1}^{n} \hat{r}_{ij}^2)^{-2} \sum_{i=1}^{n} \hat{r}_{ij}^2 Var(u_i|(x_{1i}, \ldots, x_{ki})\forall i) \\
&\overset{MLR.2-5}{=} \sigma^2 / \sum_{i=1}^{n} \hat{r}_{ij}^2 = \sigma^2/(SST_j(1 - R_j^2))
\end{aligned}$$

since $\sum_{i=1}^n \hat{r}_{ij}^2 = SST_j(1 - R_j^2)$:

$$R_j^2 = 1 - \frac{\sum_{i=1}^n \hat{r}_{ij}^2}{\sum_{i=1}^n (x_{ij} - \overline{x}_j)^2} \equiv 1 - \frac{\sum_{i=1}^n \hat{r}_{ij}^2}{SST_j}$$

Estimating error variance

Error Variance

Under assumptions MLR.1 to MLR.5 we have that

$$E[\hat{\sigma}^2 | x_1, \ldots, x_k] = \sigma^2$$

where

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

Proof of this result is omitted.

($n - k - 1 =$ #obs - #parameters = degrees of freedom)

OLS in matrix form

First write the regression model

$$y = X\beta + u$$

where

$$\underset{n\times 1}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \quad \underset{n\times(k+1)}{X} = \begin{pmatrix} x_{10} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n0} & \cdots & x_{nk} \end{pmatrix}; \quad \underset{n\times 1}{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix};$$

$$\underset{(k+1)\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$$

and $x_{i0} = 1 \, \forall i$ if the model includes a constant.

OLS in matrix form

We can, as before, obtain the OLS estimator in two ways. Either we impose that the empirical moment condition $X'(y - Xb) = 0$ and find the value of b that solves the $k + 1$ equalities, or we can write the sum of squares

$$SSR(b) = (y - Xb)'(y - Xb) = y'y - 2b'X'y + b'X'Xb$$

and solve the 1st order conditions

$$\frac{\partial SSR(b)}{\partial b} = -2X'y + 2X'Xb = 0$$

doing so gives

$$\begin{aligned} X'(y - X\hat{\beta}) &= 0 \\ X'X\hat{\beta} &= X'y \\ \hat{\beta} &= (X'X)^{-1}X'y \end{aligned}$$

OLS in matrix form

Unbiasedness follows from

$$E[\hat{\beta}|X] = E[(X'X)^{-1}X'(X\beta + u)|X] = \beta + (X'X)^{-1}X'E[u|X] = \beta$$

Note that we use the linear in parameters assumption in the first equality, and the zero conditional mean assumption in the last one. We also need that $X'X$ is invertible which is a rank condition on X (it follows from MLR.4).
The variance can also be derived

$$\begin{aligned}
Var(\hat{\beta}|X) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] \\
&= E[(X'X)^{-1}X'uu'X(X'X)^{-1}|X] \\
&= (X'X)^{-1}X'E[uu'|X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

BLUE

Gauss-Markov Theorem

Under assumption MLR.1 through MLR.5 $\hat{\beta}_j$ is the best linear unbiased estimator (BLUE) of β_j $\forall j = 0, 1, \dots, k$.

- **B**est = smallest variance
- **L**inear (in y_i): $\hat{\beta}_j = \sum_{i=1}^{n} w_i y_i$
- **U**nbiased: $E[\hat{\beta}_j] = \beta_j$
- **E**stimator: $\hat{\beta}_j = \mathit{function}(\text{data})$

OLS in matrix form: $\hat{\beta}$ is BLUE

Consider an alternative linear estimator

$$\tilde{\beta} = A'y$$

which has the following conditional expectation

$$\begin{aligned} E[\tilde{\beta}|X] &= E[A'y|X] = E[A'(X\beta + u)|X] \\ &= A'X\beta + A'E[u|X] = A'X\beta \end{aligned}$$

For $\tilde{\beta}$ to be unbiased it must be true that

$$A'X = I$$

The variance

$$var(\tilde{\beta}|X) = A'var(u|X)A = \sigma^2 A'A$$

OLS in matrix form: $\hat{\beta}$ is BLUE (cntd)

$$
\begin{aligned}
var(\tilde{\beta}|X) - var(\hat{\beta}|X) &= \sigma^2[A'A - (X'X)^{-1}] \\
&= \sigma^2[A'A - A'X(X'X)^{-1}X'A] \\
&= \sigma^2 A'[I - X(X'X)^{-1}X']A \equiv \sigma^2 A'MA \geq 0
\end{aligned}
$$

where $M = I - X(X'X)^{-1}X'$. The last inequality follows because $A'MA$ is positive semi-definite since

- $M = M'$ (symmetric)
- $M'M = M$ (idempotent)

This result carries over to any linear combination of OLS estimators since (c is a k by 1 vector):

$$var(c'\tilde{\beta}|X) - var(c'\hat{\beta}|X) = c'[var(\tilde{\beta}|X) - var(\hat{\beta}|X)]c \geq 0$$

I understand/can apply..

- MLR/OLS estimators:
  - hold many factors fixed
  - allows us to use more general functional forms
  - better prediction
  - unbiased under assumptions MLR.1-4
  - BLUE under MLR.1-5
- Partial out variables using auxiliary regression
- Omitted variable bias
- Irrelevant variables decrease the precision of OLS estimators, but do not result in bias
- R^2 measures fit. Do not pay too much attention to it (unless you are interested in prediction)