



Quantitative Methods for Economics

Tutorial 8

Katherine Eyal



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 2.5 South Africa License](https://creativecommons.org/licenses/by-nc-sa/2.5/za/).



TUTORIAL 8

27 September 2010

ECO3021S

Part A: Problems

1. The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990)¹ to study the trade-off between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u$$

where *sleep* and *totwrk* (total work) are measured in minutes per week and *educ* (years of education) and *age* are measured in years.

- (a) If adults trade off sleep for work, what is the sign of β_1 ?
- (b) What signs do you think β_2 and β_3 will have?
- (c) Using the data in Biddle and Hamermesh (1990), the estimated equation is

$$\begin{aligned}\widehat{sleep} &= 3,638.25 - 0.148 \text{ totwrk} - 11.13 \text{ educ} + 2.20 \text{ age} \\ n &= 706, R^2 = 0.113\end{aligned}$$

If someone works five more hours per week, by how many minutes is *sleep* predicted to fall? Is this a large trade-off?

- (d) Discuss the sign and magnitude of the estimated coefficient on *educ*.
 - (e) Would you say *totwrk*, *educ* and *age* explain much of the variation in *sleep*? What other factors might affect the time spent sleeping? Are these likely to be correlated with *totwrk*?
2. The following equation describes the median housing price in a community in terms of amount of pollution (*nox* for nitrous oxide) and the average number of rooms in houses in the community (*rooms*):

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{rooms} + u$$

- (a) What are the probable signs of β_1 and β_2 ? What is the interpretation of β_1 ? Explain.

¹J.E. Biddle and D.S. Hamermesh (1990), "Sleep and the Allocation of Time," *Journal of Political Economy* 98, 922-943.

- (b) Why might nox [or more precisely, $\log(nox)$] and $rooms$ be negatively correlated? If this is the case, does the simple regression of $\log(price)$ on $\log(nox)$ produce an upward or a downward biased estimator of β_1 ?
- (c) An empirical analysis yields the following results:

$$\begin{aligned}\widehat{\log(price)} &= 11.71 - 1.043 \log(nox), & n = 506, R^2 = 0.264 \\ \widehat{\log(price)} &= 9.23 - 0.718 \log(nox) + 0.306 rooms, & n = 506, R^2 = 0.514\end{aligned}$$

Is the relationship between the simple and multiple regression estimates of the elasticity of $price$ with respect to nox what you would have predicted, given your answer in part (b)? Does this mean that -0.718 is definitely closer to the true elasticity than -1.043 ?

Part B: Computer Exercises

- How do growing weather and a wine's age influence a Bordeaux wine's price? The data set WINEWEATHER1.DTA contains average 1983 prices for Bordeaux wines for the vintages from 1952 to 1980, along with data on weather conditions when each vintage was being grown. These data were part of an analysis of Bordeaux wine as an investment by economists Orley Ashenfelter, David Ashmore, and Robert LaLonde. The variables in the file are

<i>vint</i>	Vintage of the wine (i.e. its year of production)
<i>logprice</i>	Natural log of the price of Bordeaux wines relative to the price of the 1961 vintage
<i>degrees</i>	Average temperature in the growing season
<i>hrain</i>	Rainfall in the harvest season
<i>wrain</i>	Winter rainfall prior to harvest season
<i>time_sv</i>	Time from 1983 back to the wine's vintage year

- Use multiple regression to explore how growing-season temperatures, harvest-season rainfall, off-season rainfall, and the age of a wine influence the natural log of a vintage's price.
- Are the signs on the variables what you expect? Briefly explain.
- How much of the variation in vintages' prices in this sample is accounted for by these explanatory variables?
- Regress the log of price on the age of the wine and an intercept term. How do you interpret the coefficients on age in this regression and in the regression in (a)?
- Why does the variable *logprice* have the value zero for the 1961 vintage?

2. Use the data in CHARITY.DTA to answer the following questions:

(a) Estimate the equation

$$gift = \beta_0 + \beta_1 mailsyear + \beta_2 giftlast + \beta_3 propresp + u$$

by OLS. How does the R -squared compare with that from the simple regression that omits $giftlast$ and $propresp$ (you estimated this simple regression in Tutorial 6)?

- (b) Interpret the coefficient on $mailsyear$. Is it bigger or smaller than the corresponding simple regression coefficient?
- (c) Interpret the coefficient on $propresp$. Be careful to notice the units of measurement of $propresp$.
- (d) Now add the variable $avggift$ to the equation. What happens to the estimated effect of $mailsyear$?
- (e) In the equation from part (d), what has happened to the coefficient on $giftlast$? What do you think is happening?

3. Use the data in WAGE1.DTA to confirm the partialling out interpretation of the OLS estimates by explicitly doing the partialling out for Example 3.2 on page 76 of Wooldridge. This first requires regression $educ$ on $exper$ and $tenure$ and saving the residuals, $\hat{\tau}_1$. Then, regress $\log(wage)$ on $\hat{\tau}_1$. Compare the coefficient on $\hat{\tau}_1$ with the coefficient on $educ$ in the regression of $\log(wage)$ on $educ$, $exper$, and $tenure$.

TUTORIAL 8 SOLUTIONS

27 September 2010

ECO3021S

Part A: Problems

1. The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990)¹ to study the trade-off between time spent sleeping and working and to look at other factors affecting sleep:

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + u$$

where *sleep* and *totwrk* (total work) are measured in minutes per week and *educ* (years of education) and *age* are measured in years.

- (a) If adults trade off sleep for work, what is the sign of β_1 ?
- (b) What signs do you think β_2 and β_3 will have?
- (c) Using the data in Biddle and Hamermesh (1990), the estimated equation is

$$\begin{aligned}\widehat{sleep} &= 3,638.25 - 0.148 \text{ totwrk} - 11.13 \text{ educ} + 2.20 \text{ age} \\ n &= 706, R^2 = 0.113\end{aligned}$$

If someone works five more hours per week, by how many minutes is *sleep* predicted to fall? Is this a large trade-off?

- (d) Discuss the sign and magnitude of the estimated coefficient on *educ*.
- (e) Would you say *totwrk*, *educ* and *age* explain much of the variation in *sleep*? What other factors might affect the time spent sleeping? Are these likely to be correlated with *totwrk*?

SOLUTION:

- (a) If adults trade off sleep for work, more work implies less sleep (other things equal), so $\beta_1 < 0$.

¹J.E. Biddle and D.S. Hamermesh (1990), "Sleep and the Allocation of Time," *Journal of Political Economy* 98, 922-943.

- (b) The signs of β_2 and β_3 are not obvious, at least to me. One could argue that more educated people like to get more out of life, and so, other things equal, they sleep less ($\beta_2 < 0$). The relationship between sleeping and age is more complicated than this model suggests, and economists are not in the best position to judge such things.
- (c) Since *totwrk* is in minutes, we must convert five hours into minutes: $\Delta \text{totwrk} = 5(60) = 300$. Then sleep is predicted to fall by $0.148(300) = 44.4$ minutes. For a week, 45 minutes less sleep is not an overwhelming change.
- (d) More education implies less predicted time sleeping, but the effect is quite small. If we assume the difference between college and high school is four years, the college graduate sleeps about 45 minutes less per week, other things equal.
- (e) Not surprisingly, the three explanatory variables explain only about 11.3% of the variation in *sleep*. One important factor in the error term is general health. Another is marital status, and whether the person has children. Health (however we measure that), marital status, and number and ages of children would generally be correlated with *totwrk*. (For example, less healthy people would tend to work less.)
2. The following equation describes the median housing price in a community in terms of amount of pollution (*nox* for nitrous oxide) and the average number of rooms in houses in the community (*rooms*):

$$\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{rooms} + u$$

- (a) What are the possible signs of β_1 and β_2 ? What is the interpretation of β_1 ? Explain.
- (b) Why might *nox* [or more precisely, $\log(\text{nox})$] and *rooms* be negatively correlated? If this is the case, does the simple regression of $\log(\text{price})$ on $\log(\text{nox})$ produce an upward or a downward biased estimator of β_1 ?
- (c) An empirical analysis yields the following results:

$$\begin{aligned} \widehat{\log(\text{price})} &= 11.71 - 1.043 \log(\text{nox}), & n = 506, R^2 = 0.264 \\ \widehat{\log(\text{price})} &= 9.23 - 0.718 \log(\text{nox}) + 0.306 \text{rooms}, & n = 506, R^2 = 0.514 \end{aligned}$$

Is the relationship between the simple and multiple regression estimates of the elasticity of *price* with respect to *nox* what you would have predicted, given your answer in part (b)? Does this mean that -0.718 is definitely closer to the true elasticity than -1.043 ?

SOLUTION:

- (a) $\beta_1 < 0$ because more pollution can be expected to lower housing values; note that β_1 is the elasticity of price with respect to nox . β_2 is probably positive because $rooms$ roughly measures the size of a house. (However, it does not allow us to distinguish homes where each room is large from homes where each room is small.)
- (b) If we assume that $rooms$ increases with quality of the home, then $\log(nox)$ and $rooms$ are negatively correlated when poorer neighborhoods have more pollution, something that is often true. We can use Table 3.2 on page 91 of Wooldridge to determine the direction of the bias. If $\beta_2 > 0$ and $\text{Corr}(x_1, x_2) < 0$, the simple regression estimator $\tilde{\beta}_1$ has a downward bias. But because $\beta_1 < 0$, this means that the simple regression, on average, overstates the importance of pollution. ($E(\tilde{\beta}_1)$ is more negative than β_1 .)
- (c) This is what we expect from the typical sample based on our analysis in part (b). The simple regression estimate, -1.043 , is more negative (larger in magnitude) than the multiple regression estimate, -0.718 . As those estimates are only for one sample, we can never know which is closer to β_1 . But if this is a “typical” sample, β_1 is closer to -0.718 .

Part B: Computer Exercises

1. How do growing weather and a wine’s age influence a Bordeaux wine’s price? The data set WINEWEATHER1.DTA contains average 1983 prices for Bordeaux wines for the vintages from 1952 to 1980, along with data on weather conditions when each vintage was being grown. These data were part of an analysis of Bordeaux wine as an investment by economists Orley Ashenfelter, David Ashmore, and Robert LaLonde. The variables in the file are

<i>vint</i>	Vintage of the wine (i.e. its year of production)
<i>logprice</i>	Natural log of the price of Bordeaux wines relative to the price of the 1961 vintage
<i>degrees</i>	Average temperature in the growing season
<i>hrain</i>	Rainfall in the harvest season
<i>wrain</i>	Winter rainfall prior to harvest season
<i>time_sv</i>	Time from 1983 back to the wine’s vintage year

- (a) Use multiple regression to explore how growing-season temperatures, harvest-season rainfall, off-season rainfall, and the age of a wine influence the natural log of a vintage’s price.

- (b) Are the signs on the variables what you expect? Briefly explain.
- (c) How much of the variation in vintages' prices in this sample is accounted for by these explanatory variables?
- (d) Regress the log of price on the age of the wine and an intercept term. How do you interpret the coefficients on age in this regression and in the regression in (a)?
- (e) Why does the variable *logprice* have the value zero for the 1961 vintage?

SOLUTION:

- (a) Command: `reg logprice wrain degrees hrain time_sv`

Model	8.66443586	4	2.16610897	Prob > F	0
Residual	1.80582883	22	.082083129	R-squared	0.8275
				Adj R-squared	0.7962
Total	10.4702647	26	.402702488	Root MSE	0.2865

<i>logprice</i>	Coef.	Std. Err	t	P>t	95% Conf. Interval	
<i>wrain</i>	0.0011668	.000482	(2.42)	2.40%	0.0001671	0.0021665
<i>degrees</i>	0.6163926	.0951755	(6.48)	0.00%	0.4190107	0.8137745
<i>hrain</i>	-0.0038606	.0008075	(-4.78)	0.00%	-0.0055353	-0.0021858
<i>time_sv</i>	0.0238474	.0071667	(3.33)	0.30%	0.0089846	0.0387103
<i>_cons</i>	-12.14534	1.688103	(-7.19)	0.00%	-15.64625	-8.644426

wrain: If winter rain increases by 100ml then the average relative price of Bordeaux wines increase by 11.66%

degrees: If the average temperature increases by 1 deg centigrade then average relative price of Bordeaux wines increases by 62%

hrain: If harvest rainfall increases by 100ml then average relative price of Bordeaux wines decreases by 38.6%

time_sv : This variable gives the age of the wine. If the age of the wine increases by 1 year then average relative price of Bordeaux wines increases by 2.4%

Note that all the variables are significant at the 1% level except *wrain* which is significant at the 5% level.

- (b) The signs are as expected one would expect winter rain fall, warm weather and the vintage to increase the value of the average relative price. Summer rain fall is likely to have a negative effect on crop yields and thus the relative price (given that it is uncharacteristic of most wine growing regions to have summer rain fall). The size of the coefficient on *degrees* is larger than the others, but one must consider that *degrees* is an average which probably has a low degree of variation from season to season.

- (c) Look at the R^2 : The explanatory variables explain roughly 83% of the variation in vintages relative prices.
- (d) Command: `reg logprice time_sv`

<i>logprice</i>	Coef.	Std. Err.	t	P>t
<i>time_sv</i>	0.0354296	0.0136625	2.59	0.016
<i>_cons</i>	-2.025199	0.2472287	-8.19	0

Note that one can use *time_sv* or *vint* to indicate the age of the wine (these two variables are perfectly collinear). Using *vint* will result in the signs being reversed (test this). Here the premium for an additional year of the vintage is now higher at 3.5% per year. The difference is now we do not control for all other factors affecting the average relative price as we did in the multiple regression, and they are thus in the error.

Note this leads to an upward bias, which is quite significant yet it is unclear why *time_sv* would be correlated with any of the other above explanatory variables should they be in the error but the “other” explanatory variables are certainly correlated with y . The most important difference is in the multiple regression we have the effect of an additional vintage year on average relative price while controlling for winter and harvest rain fall as well as the average temperature.

- (e) The price of the 1961 vintage relative to the price of the 1961 vintage is 1, and the log of 1 is zero.

2. Use the data in CHARITY.DTA to answer the following questions:

- (a) Estimate the equation

$$gift = \beta_0 + \beta_1 mailsyear + \beta_2 giftlast + \beta_3 propresp + u$$

by OLS. How does the R -squared compare with that from the simple regression that omits *giftlast* and *propresp* (you estimated this simple regression in Tutorial 6)?

- (b) Interpret the coefficient on *mailsyear*. Is it bigger or smaller than the corresponding simple regression coefficient?
- (c) Interpret the coefficient on *propresp*. Be careful to notice the units of measurement of *propresp*.
- (d) Now add the variable *avggift* to the equation. What happens to the estimated effect of *mailsyear*?
- (e) In the equation from part (d), what has happened to the coefficient on *giftlast*? What do you think is happening?

SOLUTION:

- (a) The estimated equation is

$$\begin{aligned}\widehat{gift} &= -4.55 + 2.17 \text{ mailsyear} + .0059 \text{ giftlast} + 15.36 \text{ propresp} \\ n &= 4,268, \quad R^2 = .0834\end{aligned}$$

The R -squared is now about .083, compared with about .014 for the simple regression case. Therefore, the variables *giftlast* and *propresp* help to explain significantly more variation in gifts in the sample (although still just over eight percent).

- (b) Holding *giftlast* and *propresp* fixed, one more mailing per year is estimated to increase *gifts* by 2.17 guilders. The simple regression estimate is 2.65, so the multiple regression estimate is somewhat smaller. Remember, the simple regression estimate holds no other factors fixed.
- (c) Because *propresp* is a proportion, it makes little sense to increase it by one. Such an increase can happen only if *propresp* goes from zero to one. Instead, consider a .10 increase in *propresp*, which means a 10 percentage point increase. Then, gift is estimated to be $15.36(.1) \approx 1.54$ guilders higher.
- (d) The estimated equation is

$$\begin{aligned}\widehat{gift} &= -7.33 + 1.20 \text{ mailsyear} - .261 \text{ giftlast} + 16.20 \text{ propresp} + .527 \text{ avggift} \\ n &= 4,268, \quad R^2 = .2005\end{aligned}$$

After controlling for the average past gift level, the effect of mailings becomes even smaller: 1.20 guilders, or less than half the effect estimated by simple regression.

- (e) After controlling for the average of past gifts – which we can view as measuring the “typical” generosity of the person and is positively related to the current gift level – we find that the current gift amount is negatively related to the most recent gift. A negative relationship makes some sense, as people might follow a large donation with a smaller one.
3. Use the data in WAGE1.DTA to confirm the partialling out interpretation of the OLS estimates by explicitly doing the partialling out for Example 3.2 on page 76 of Wooldridge. This first requires regression *educ* on *exper* and *tenure* and saving the residuals, \hat{r}_1 . Then, regress $\log(\text{wage})$ on \hat{r}_1 . Compare the coefficient on \hat{r}_1 with the coefficient on *educ* in the regression of $\log(\text{wage})$ on *educ*, *exper*, and *tenure*.

SOLUTION:

The regression of *educ* on *exper* and *tenure* yields

$$\begin{aligned}educ &= 13.57 - .074 \textit{exper} + .048 \textit{tenure} + \hat{r}_1 \\n &= 526, \quad R^2 = .101.\end{aligned}$$

Now, when we regress $\log(\textit{wage})$ on \hat{r}_1 we obtain

$$\begin{aligned}\log(\textit{wage}) &= 1.62 + .092 \hat{r}_1 \\n &= 526, \quad R^2 = .207\end{aligned}$$

As expected, the coefficient on \hat{r}_1 in the second regression is identical to the coefficient on *educ* in equation (3.19) on page 76 of Wooldridge. Notice that the *R*-squared from the above regression is less than that in (3.19). In effect, the regression of $\log(\textit{wage})$ on \hat{r}_1 explains $\log(\textit{wage})$ using only the part of *educ* that is uncorrelated with *exper* and *tenure*; separate effects of *exper* and *tenure* are not included.

TUTORIAL 2, QUESTION 8

There are basically two steps to this:

1. Show that $AI = A$ and $IA = A$ for any (3×3) matrix A

$$\text{and } I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

2. Show that there is no other matrix, J , that satisfies

$$JA = A, \quad AJ = A \quad \text{for any } A$$

('other' meaning $J \neq I$)

Step 2

$AI = A$:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} (a_{11} \cdot 1 + a_{12} \cdot 0 + a_{13} \cdot 0) & \text{etc.} \\ (a_{21} \cdot 1 + a_{22} \cdot 0 + a_{23} \cdot 0) & \text{etc.} \\ \text{etc.} & \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = A$$

and similarly for $IA = A$

Step 2

We show that any J satisfying $JA = AJ = A$ equals I .

Assume there is some J that satisfies $JA = A$, for any A

Let $A = I$ (we can choose any matrix after all...)

then $JI = I$

But we know $AI = A$ for any A (see step 1)

So we must have $JI = J$

$$\text{or } I = JI = J$$

i.e. There is no matrix not equal to I that satisfies the requirements for an identity matrix. [Note: More formally written, this would be a 'proof by contradiction']