

# Econometrics Problem Set 1

Cheryl Lim

October 4th 2024

## EXERCISE 1.A - Sleep and work

The data set SLEEP75.dta from Biddle and Hamermesh (1990) contains data on the number of minutes per week spend sleeping (sleep) and the number of minutes spend in paid work (totwrk).

When the following simple regression was estimated using data on  $n = 706$  individuals:

$$\text{sleep} = \beta_0 + \beta_1 \text{totwrk} + u$$

the estimated slope coefficient was  $-0.15$ , while the estimated intercept was  $3586.38$ .

**1. Interpretation of slope coefficient:** The slope coefficient  $\beta_1 = -0.15$  is associated with a negative relationship between sleep and paid work, where an additional minute per week spent working in a paid job is associated with a decrease in the amount of time spent on sleep by  $0.15$  minutes (or  $9$  seconds) per week, ceteris paribus.

**2. Prediction of sleep for someone who does not work in a paid job:** For a person who does not work ( $\text{totwrk}=0$ ) in a paid job the whole week, the predicted number of minutes of sleep is:

$$\text{sleep} = \beta_0 + \beta_1 \times (0) = 3586.38 \text{ minutes per week}$$

The intercept tells us that on average, someone who doesn't work in a paid job spends  $3586.38$  minutes per week sleeping, which is approximately  $8.55$  hours per night. We notice that it is at the higher end of CDC's recommended amount of sleep for adults.

**3. Prediction for a sleeping time of 8 hours:** To predict for a sleeping time of  $8$  hours/per night, we set  $\text{sleep} = 3360$  since  $8 \text{ hours per night} = 3360 \text{ minutes per week}$ , :

$$\begin{aligned} \text{sleep} &= \beta_0 + \beta_1 \text{totwrk} + u \\ \Rightarrow 3360 &= 3586.38 - 0.15 * \text{totwrk} \\ \Leftrightarrow \text{totwork} &= 1509.2 \text{ minutes per week} \approx 25.15 \text{ hours per week} \end{aligned}$$

**4. Factors contained in  $u$ :**  $u$  is the error term which captures all the observable and unobservable factors that affect sleep but are not included in the regression model (different from minutes spent in paid work). These may include, for example, an individual's health, age, gender or household composition. Some of these factors are likely to be also correlated with the amount of time spent in paid work.

An example of such factors could be time spent on other activities apart from sleep & work. These factors are likely to be either positively/negatively correlated with the amount of time spend working since time is finite - everyone has a limited number of hours everyday to spend on different activities so there will always be a trade-off with regard to the amount of time spent on each activity. For instance, having children or not has an impact on both the time spent at work and the time spent sleeping.

**5. Simple regression sufficient to uncover the ceteris paribus effect of time spent sleepin and working?** No, a simple regression analysis will not uncover the ceteris paribus effect of time spent working.

This is because there are likely omitted variables affecting both sleep and work that are not included in the simple regression (stress, number of children, number of jobs, health, etc.). Not including them could lead to omitted variable bias when running the regression and, for example, lead to overestimate or underestimate the true effect of work on sleep. Running a multivariate regression will help control for more factors and obtain a more accurate estimate of the effect of work on sleep.

## EXERCISE 1.B - impact of free school meals program

Imagine you work in a poor country where many children drop out of school, or are often absent. You are interested in knowing the impact of a program providing free school meals on learning outcomes. As you are worried about omitted variable bias, you randomly provide school meals in some schools, and not in others. When you obtain the results of the end-of-the-year exam you find that schools with free meals have lower average test results?

**1. Explanation for lower average test results:** By providing some schools with meals and others without, selection bias could have been introduced. Therefore, if the schools that happened to be selected already contained students with lower initial learning outcomes, the lower test results could simply be a reflection of their current level of learning rather than the effect of the free school meals program itself.

There could also be spillover effects where students who start attending for free meals happen to be weaker students that negatively impact their peers (this, however, would violate our iid assumption).

Attrition bias could have occurred during the course of the study.

Composition effect and nature of the sampling distribution could also have affected the results obtained.

**2. Policy implications of test results:** We should not use these results to advise against free meals, as the program might have other positive outcomes. First, free meals could have positive effects that are not immediately reflected in school performances, such as better health, more happiness. Second, these free meals could improve attendance and long-term learning outcomes beyond the one year, that the study's time frame didn't account for. More research/analysis should be done to understand the negative correlation.

**3. Other types of data that may be useful in determining the impact of the program on children's learning:**

- Initial/baseline test scores, in order to compare the performances before and after the program.
- Socioeconomic background of each student : parents' education, parents' income, size of the household etc.
- Attendance data, to see if free meals were a real incentive to attend school.
- Characteristics of the school : size of the class, number of teachers, school type, location etc.

## EXERCISE 2 - GNI and business

The data set busind.dta contains information on Gross National Income (GNI) per capita and the number of days to open a business and to enforce a contract in a sample of 155 countries. It was extracted from the "Doing Business" dataset, a dataset collected by the World Bank based on expert opinions in each country.<sup>2</sup> The variable gnipc measures GNI per capita in \$. The variable daysopen measures the average number of days needed to open a business in that country, and daysenforce measures the average number of days needed to enforce a given type of contract.

**1. Means of important variables:**

AverageGNIPerCapita	AverageDaysToOpen	AverageDaysToEnforce
7202.194	47.51613	375.7434

## 2. Opening a business:

There are 2 countries which require less than 5 days to open a business :Australia, Canada.

In our sample, the maximum number of days needed to open a business is 203.

There is one country which requires more than 200 days to open a business :Haiti.

## 3. Regression Model 1:

$$\text{Regression model: } gnipc = \beta_0 + \beta_1 * daysopen + u$$

Table 2: GNI per capita and days needed to open a business

	<i>Dependent variable:</i>
	gnipc
Days for opening	-110.855*** (22.966)
Constant	12,469.580*** (1,391.695)
Observations	155
R <sup>2</sup>	0.132
Adjusted R <sup>2</sup>	0.126
Residual Std. Error	10,753.240 (df = 153)
F Statistic	23.300*** (df = 1; 153)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

- Estimate of  $\beta_1 = -110.85$  : we observe that ceteris paribus, a day increase in the number of days needed to open a business is associated to a decrease of 110\$ of GNI per capita. This results are significant at the 1% level. The negative sign was expected as less efficient processes to open a business might reflect bureaucratic inefficiencies, which can be correlated with lower economic performance.
- Estimate of  $\beta_0$  : the intercept tells us that in a hypothetical country where opening a business takes 0 day, the average GNI per capita is expected to be equal to 12469.58. As we are interested in GNI per capita, a positive value was expected.

**4. Error term  $u$ :**  $u$  is the error term which captures all the observable and unobservable factors that affect GNI per capita but are not included in the regression model.

An example of such factors could be political conditions, infrastructures, institutional quality, business taxation, human capital... Some of these are likely to be correlated with the number of days to open a business. For instance, the quality (and/or presence) of infrastructure may impede the process of opening a business if business owners have to source/provide their own infrastructure. Efficiency of the administrative system (bureaucracy) can also be correlated with both the number of days needed to open a business and the GNI per capita.

## 5. Predicted GNI:

- Predicted GNI per capita for a country where it takes 5 days to open a business: 11915.31
- Predicted GNI per capita for a country where it takes 200 days to open a business : -9701.376

$$\text{Given: } gnipc = \beta_0 + \beta_1 \times \text{daysopen} + u$$

- Manual calculation of predicted GNI for a country where it takes 5 days to open a business:  $gnipc = 12469.58 - (110.85 \times 5) = 11915.33$
- Manual calculation of predicted GNI for a country where it takes 200 days to open a business:  $gnipc = 12469.58 - (110.85 \times 200) = -9700.42$

The obtained level of income for a country where it takes 5 days to open a business seems reasonable, especially if it were in a middle to high-income economy. However, we know that Canada was among the countries for which it takes less than 5 days to open the business, and in 2021, its GNI per capita was \$53,310. The prediction we found seems low in that case, indicating that the number of days it takes to open a business is not sufficient to perfectly predict GNI per capita of a country.

The obtained level of income for a country where it takes 200 days to open a business seems unreasonable because GNI per capita should not be negative. The negative income could simply be a result of using a linear model.

## 6. Regression model 2:

Table 3: GNI per capita and days needed to enforce a contract

<i>Dependent variable:</i>	
	gnipc
Days for enforcement	-27.085*** (4.676)
Constant	17,328.740*** (1,950.064)
Observations	152
R <sup>2</sup>	0.183
Adjusted R <sup>2</sup>	0.177
Residual Std. Error	10,429.080 (df = 150)
F Statistic	33.549*** (df = 1; 150)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Interpretation of  $\beta_1 = -27.085$ : we observe that ceteris paribus, a day increase in the number of days needed to enforce a contract is associated to a decrease of 27.085\$ of GNI per capita. This results are significant at the 1% level. This makes sense as delays can impede economic performances of the business, leading to higher costs, less investments, less economic activity, which, in the end, could be associated with lower GNI per capita.

## 7. Comparisons of models 1 and 2:

- *Adjusted R<sup>2</sup>* for model 1 = 0.1321599
- *Adjusted R<sup>2</sup>* for model 2 = 0.1827794

A higher  $R^2$  indicates that the model explains more of the variations in GNI per capita across countries (indicates a better fit). Therefore, we see that the model using the variable *daysenforce* explains more of the variation in GNI per capita, and so, that the duration for enforcing contracts is more strongly correlated with income per capita than the duration to open a business.

Table 4:  $\log(\text{GNI per capita})$  and days needed to open a business

	<i>Dependent variable:</i>
	$\log(\text{gnipc})$
Days for opening	-0.016*** (0.003)
Constant	8.450*** (0.192)
Observations	155
R <sup>2</sup>	0.143
Adjusted R <sup>2</sup>	0.137
Residual Std. Error	1.483 (df = 153)
F Statistic	25.516*** (df = 1; 153)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

### 8. Regression model 3:

We observe that *ceteris paribus*, a day increase in the number of days needed to open a business is associated to a decrease of 1.6% of GNI per capita. This results are significant at the 1% level. This suggests that countries where it takes longer to open a business tend to have lower GNI per capita. As we mentioned in (3), long duration to open a business might reflect bureaucratic inefficiencies, which can be correlated with lower economic performance.

**9. Policy implications:** Results from the linear and log regression models allow us to conclude that policies aimed at reducing the number of days for opening a business in certain developing countries can be desirable, particularly for increasing a country's GNI. This is evidenced by the estimates from both models, which show that each additional day needed to open a business is associated with a decrease in GNI per capita of approximately \$110.85 (or 1.6% in the log model), along with both models being statistically significant.

Nevertheless, these regressions are likely to suffer from a omitted variable bias: numerous other factors could influence GNI per capita, such as political stability, infrastructure quality, and levels of human capital. Consequently, a policy focused exclusively on reducing the number of days to open a business may not be the most effective solution, as other interventions could as other might be more suitable for overall development.

Moreover, such policies might not have any effects if it is not accompanied by more structural changes in the administrative and economic systems of the countries.

Finally, it seems necessary to take into account all the societal implications of such policies beyond merely increasing GNI. A reduction in the time required to start a business could for example strengthen business owners at the expense of other segments of the population, potentially exacerbating inequalities.

**10. Missing countries:** It is important to account for the fact that the dataset only contains a sample of 155 countries. Overall, it is not clear if it was constructed from a random sample of countries : maybe, for example, countries with better and more available data are more represented in our sample than others. Therefore, there is a weak external validity : the generalization of the results cannot be reliable and the conclusions do not accurately reflect the dynamics in countries not included in the dataset.