MOEC0021 - Empirical Methods

Group BlancSchneiderMazidi

Fabienne Blanc (15-732-142)

Flavio Schneider (15-716-202)

Manuel Mazidi (15-704-984)

Course

Empirical Methods
Prof. Greg Crawford
University of Zurich

Submitted on November 19, 2018

1 Pencil and Paper Questions

Exercise 1 – Coefficients Interpretation.

(a) We estimate the following model:

$$consumption_i = \beta_0 + \beta_1 * income_i + \epsilon_i = \ 5{,}800.441 + 0.267 \ * income_i + \epsilon_i$$

The value of the coefficient on income is 0.267. So, based on this regression, an additional USD in income corresponds to an increase of consumption by 0.267 USD. In a more economical way this can be interpreted, that an average household spends roughly 27% of its income additional to some "fixed costs" of about 5,800 USD for goods which fall under the specified consumption category.

(b) We estimate the following model:

$$\begin{aligned} consumption_i &= \beta_0 + \beta_1 * income_i + \beta_2 * fam\_size_i + \epsilon_i \\ &= \ 4{,}429.220 + 0.254 \ * income_i + 625.431 * fam\_size_i + \epsilon_i \end{aligned}$$

The value of the coefficient on income is now lower at 0.254. So, based on this new regression, an additional USD in income corresponds to an increase of consumption by 0.254 USD, assuming everything else constant. As for family size we estimate that an average household spends around 625 USD per additional family member, assuming everything else constant.

(c) We estimate the following model:

$$\begin{aligned} consumption_i &= \beta_0 + \beta_1 * income_i + \beta_2 * fam\_size_i + \beta_3 * house_i + \epsilon_i \\ &= \ 4{,}429.220 + 0.254 \ * income_i + 625.431 * fam\_size_i + 1{,}395.781 * house_i + \epsilon_i \end{aligned}$$

The value of the coefficient on income does not change compared to the model before and stays at 0.254. So, based on this regression, an additional USD in income corresponds to an increase of consumption by 0.254 USD, assuming everything else constant.

(d) As we can observe the interpretation of the coefficient on income stays roughly the same. The less variables we include in our model the stronger the interpretation gets, as for example in model (1) it is the only variable we can interpret. On another note we can also see, that the constant term is decreasing with every additional variable in the models (2) and (3). This may be an indication that these variables do have an impact on the dependent variable and shouldn't be omitted. According to this argumentation the estimated coefficient on income is more precise if we include these variables. Therefore, we should try to get the most precise interpretation we can achieve and go for model (3).

Exercise 2 – Omitted Variable Bias.

We suppose that the true data generating process for a student's salary in their first job after graduation with a Master's degree is:

$$Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon_i$$

Where: Y_i = starting salary of individual i, X_{1i} = individual i's grade point average (GPA) in its Master's coursework, X_{2i} = a dummy for whether an individual i's Master's degree was in economics or finance (vs. history or literature or other). We estimate the following model:

$$Y_i = \alpha_0 + \alpha_1 * X_{1i} + \epsilon_i$$

We assume that the standard CLRM assumptions hold.

(a) The estimated coefficient of α_1 ($\hat{\alpha}_1$) can be written down as follows:

$$\hat{\alpha}_1 = (X_{1i}{}'X_{1i})^{-1}X_{1i}{}'Y_i$$

Now, the expected value of $\hat{\alpha}_1$ given X_{1i} (i.e. $E(\hat{\alpha}_1|X_{1i})$) can be calculated as follows:

$$E(\hat{\alpha}_1|X_{1i}) = E[(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'Y_i]$$

For Y_i, we can plug in the true model: $Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon_i$

$$E(\hat{\alpha}_1|X_{1i}) = E[(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(\beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \epsilon_i)]$$

Where:

- $E[(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(\beta_0)] = 0$,
- $E[(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(\beta_1 * X_{1i})] = \beta_1$,
- $E[(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(\epsilon_i)] = 0$ (Because of Assumption 2)

Therefore:

$$E(\hat{\alpha}_1|X_{1i}) = \beta_1 + E[\beta_2(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(X_{2i})]$$

$$E(\hat{\alpha}_1|X_{1i}) = \beta_1 + \beta_2(X_{1i}{}'X_{1i})^{-1}X_{1i}{}'(X_{2i})$$

$$\boldsymbol{E(\hat{\alpha}_1|X_{1i}) = \beta_1 + \beta_2 * \hat{\beta}_{X_{2i\_on\_X_{1i}}}}$$

Where: $\beta_2 * \hat{\beta}_{X_{2i\_on\_X_{1i}}}$ is the OLS coefficient from the regression of X_{2i} on X_{1i}.

(b) The estimated coefficient $\hat{\alpha}_1$ is not biased if $E(\hat{\alpha}_1|X_{1i}) = \beta_1$. However, we found in (a) that $E(\hat{\alpha}_1|X_{1i}) = \beta_1 + \beta_2 * \hat{\beta}_{X_{2i\_on\_X_{1i}}}$ whereas the additional term $\beta_2 * \hat{\beta}_{X_{2i\_on\_X_{1i}}}$ is likely to be $\neq 0$ ($\beta_2 = 0$ would imply that no relevant variable is omitted; $\hat{\beta}_{X_{2i\_on\_X_{1i}}} = 0$ would imply that each of the elements of X_{1i} is uncorrelated with X_{2i}). Consequently, we suppose that the estimated coefficient $\boldsymbol{\hat{\alpha}_1}$ **is likely to be biased**. In this case, bias is caused by the omission of a relevant variable (X_{2i}) in the estimated model.

(c) In order to assess the sign of the bias, we make the following assumption (reason: this considers the "weakest form" of bias): Only one element of X_{1i} is correlated with X_2. Hence, $X_{1,K}{}'X_2 = \sigma_{K,X_2}, X_{1,k}{}'X_2 = 0$ for $\forall k = 1, \dots, K-1$.

$$\hat{\alpha}_k = \left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}Y_k$$

$$E(\hat{\alpha}_k) = E[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}Y_k]$$

For Y_i, we again plug in the true model: $Y_i = \beta_0 + \beta_1 * X_{1,i} + \beta_2 * X_2 + \epsilon_i$

$$E(\hat{\alpha}_k) = E[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}(\beta_0 + \beta_1 * X_{1,k} + \beta_2 * X_2 + \epsilon_k)]$$

Where:

- $E\left[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}(\beta_0)\right] = 0,$
- $E\left[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}'M_{1,-k}(\beta_1 * X_{1,k})\right] = \beta_1,$
- $E\left[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}(\epsilon_k)\right] = 0$ (by Assumption 2)

Therefore:

$$E(\hat{\alpha}_k) = E[\left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}(\beta_1 * X_{1,k} + \beta_2 * X_2)]$$

$$\boldsymbol{E(\hat{\alpha}_k) = \beta_1 + \beta_2 * \left(X_{1,k}{}'M_{1,-k}X_{1,k}\right)^{-1} X_{1,k}{}'M_{1,-k}X_2}$$

Interpretation – two terms determine the sign of the bias:

- β_2 represents the impact of the omitted variable (X_{2i}) on Y_i, i.e. the impact of having a Master's degree in economics or finance (vs. history or other) on the starting salary.

➔ We assume – based on recent studies – that salaries in the economics and finance industry a higher compared to other industries. Therefore, β_2 should be positive.

- $(X_{1,k}{}'M_{1,-k}X_{1,k})^{-1}X_{1,k}{}'M_{1,-k}X_2 = (X_k^{*}{}'X_k^{*})^{-1}X_k^{*}{}'X_2^{*}$ (where X_k^{*} and X_2^{*} are the residuals from the regression of each variable on $X_{1,-k}$). This term represents the correlation between $X_{1,k}$ and X_2, while controlling for the other elements in $X_{1,k}$, i.e. $X_{1,-k}$. In other words, this term measures the correlation between an individual's grade point average (GPA) in its Master's coursework and the fact whether an individual's Master's degree was in economics or finance (vs. history or literature or other).

  ➔ We assume that this correlation is nearly zero. We find it plausible that grades in economics and finance are not systematically higher or lower than grades in other field of studies such as history or literature. Hence, the specific field of study would not impact an individual's grade point average (GPA). Vice versa, there is no meaningful interpretation (the obtained GPA cannot have an impact on the study field).

➔ We conclude that the bias is likely to be **positive**.

(d) Now, we suppose that the true data generating process includes another variable (X_{3i}), which measures the time spent per week by an individual i on extracurricular activities. As we are able to include this variable in our estimated model, we formulate the following models:

True model: $Y_i = \beta_0 + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \beta_3 * X_{3i} + \epsilon_i$

Estimated model: $Y_i = \alpha_0 + \alpha_1 * X_{1i} + \alpha_3 * X_{3i} + \varepsilon_i$

Going back to (c), we find that the addition of this new variable has the following impact:

- In (c), we claim that the model is biased. By adding a new variable, the model does not suddenly become unbiased, as it still omits the impact of the study field (X_{2i}). Hence, the model is still **biased**.

- Further, in (c), we claim that the model is positively biased. Now, we assume that the new variable (X_{3i}) has a positive impact on the starting salary (as companies often seem to support and value extracurricular activities). Therefore, the estimated model in (c) did omit another relevant variable (with a positive correlation to Y_i), which is no longer omitted in the new estimated model. Consequently, the **bias of this new estimated model is smaller – but still positive – compared to the estimated model (c)**.

Exercise 3 – Measurement Error in y.

For this exercise, we suppose that the true model is: $y_i^* = x_i'\beta + \epsilon_i^*$ with $E(\epsilon_i^*|x_i) = 0$ and $V(\epsilon_i^*|x_i) = \sigma_*^2$ (as in the CLRM). There is a measurement error in y_i: $y_i = y_i^* + \eta_i$ with $\eta_i \sim (0, \sigma_\eta^2)$. Hence, the estimated model is: $y_i = x_i'\beta + \epsilon_i$. Further, we assume that this is "classical measurement error", i.e., it is uncorrelated with everything: $E(\eta_i|x_i) = 0$ and $E(\eta_i|\epsilon_i^*) = 0$.

(a) Because of the measurement error, the y_i we see (y_i) is equal to the true value (y_i^*) plus an error term (η_i). Hence, the estimated model is: $y_i = y_i^* + \eta_i = x_i'\beta + \epsilon_i$. Knowing that, we can transform the true model as follows:

$$y_i^* = x_i'\beta + \epsilon_i^*$$

Adding η_i on the left- and right-hand side:

$$y_i^* + \eta_i = x_i'\beta + \epsilon_i^* + \eta_i$$

$$y_i = x_i'\beta + \epsilon_i^* + \eta_i$$

$$y_i = x_i'\beta + \epsilon_i$$

Where $\epsilon_i = \epsilon_i^* + \eta_i$ is a "composite error" of the true error (ϵ_i^*) & the measurement error (η_i).

➔ Mean: $E(\epsilon_i) = E(\epsilon_i^* + \eta_i)$

$$E(\epsilon_i) = E(\epsilon_i^*) + E(\eta_i)$$

Supposing that assumption 2 ("Mean-Zero Error") holds, $E(\epsilon_i^*) = 0$. Similarly, $E(\eta_i) = 0$.

$$\boldsymbol{E(\epsilon_i) = 0}$$

➔ Variance: $V(\epsilon_i) = V(\epsilon_i^* + \eta_i)$

$$V(\epsilon_i) = V(\epsilon_i^*) + V(\eta_i) + 2Cov(\epsilon_i^*; \eta_i)$$

Given our assumption made before ("classical measurement error"), $2Cov(\epsilon_i^*; \eta_i) = 0$.

Supposing that assumption 3 ("Homoskedasticity") holds, $V(\epsilon_i^*) = \sigma^2$.

$$\boldsymbol{V(\epsilon_i) = \sigma^2 + \sigma_{\eta_i}^2}$$

(b) Is $\hat{\beta}$ biased in this case?

$$\hat{\beta} = \beta + (X'X)^{-1}X'\epsilon$$

$$= \beta + \frac{X'\epsilon}{X'X}$$

$$= \beta + \frac{\frac{1}{N}X'\epsilon}{\frac{1}{N}X'X}$$

Probability limit:

$$\hat{\beta} = p\lim_{n\to\infty}(\beta + \frac{\frac{1}{N}X'\epsilon}{\frac{1}{N}X'X})$$

Assuming only one explanatory variable (x_i):

$$\hat{\beta} = \beta + \frac{cov(\epsilon_i; x_i)}{var(\epsilon_i) * var(x_i)}$$

Where $cov(\epsilon_i; x_i) = cov(\epsilon_i^* + \eta_i; x_i) = cov(\epsilon_i^*; x_i) + cov(\eta_i; x_i) = 0 + 0 = 0$ because of Assumption 2 ($E(\epsilon_i^*|x_i) = 0$ implying $cov(\epsilon_i^*; x_i) = 0$) and the "classical measurement error" ($E(\eta_i|x_i) = 0$ implying $cov(\eta_i; x_i) = 0$).

Therefore:

$$\hat{\beta} = \beta + \frac{0}{var(\epsilon_i) * var(x_i)}$$

$$\boldsymbol{\hat{\beta} = \beta}$$

We conclude that **$\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$ even when there is a measurement error in $\boldsymbol{y_i}$**.

2 Computer Questions

Exercise 1 – Dealing with Measurement Error.

(a) We expect to have a large measurement error in the variable *corruptionun* simply because there is no standardized method how the extent of corruption of a single country can be measurement. Furthermore, corruption can only be measured when it is visible to the organization that wants to measure the extend of corruption. It is very likely therefore that many cases of corruption are not taken account of as corruption usually is not reported when it is not uncovered and prosecuted actively.

The measurement error for child mortality is most likely smaller than for corruption as these usually have to be reported to officials and the definition of child mortality is very strict in contrary to corruption. Nevertheless, there might be a small measurement error as not all child deaths are reported.

(b) i.

```
================================================
                         Dependent variable:
                     ---------------------------
                             mortalityun
------------------------------------------------
corruptionun                  0.626***
                               (0.083)

Constant                       0.00000
                               (0.083)

------------------------------------------------
Observations                     90
R2                              0.392
Adjusted R2                     0.385
Residual Std. Error        0.784 (df = 88)
F Statistic            56.685*** (df = 1; 88)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Our OLS-estimate is $\hat{\beta} = 0.626$. The one-sided hypothesis test with $H_0\colon \hat{\beta} \leq 0$ is calculated in the following way.

$$t_{statistic} = \frac{\hat{\beta} - \beta_0}{\sigma_{\hat{\beta}}} = \frac{0.626}{0.083} = 7.53$$

We calculated a t-statistics of 7.53 which exceeds all customary one-sided critical values. Thus, H0 is rejected in favour of the alternative hypothesis that is $H_A\colon \hat{\beta} > 0$.

ii.

The coefficient $\hat{\beta}$ is interpreted in the following way: The beta estimate of 0.626 is statistically significant and indicates that higher corruption leads to higher child mortality (positive sign). If *corruptionun* increases by one unit, then on average we expect an increase in *mortalityun* of 0.626 in this simple regression. This number does not tell us on the actual impact in child mortality as we do not know exactly how the index was constructed, however, we do know that the maxima of *mortalityun* is roughly 3.2 in our dataset. Therefore, the relative increase of 0.626 is rather high and thus of great economic relevance.

iii.

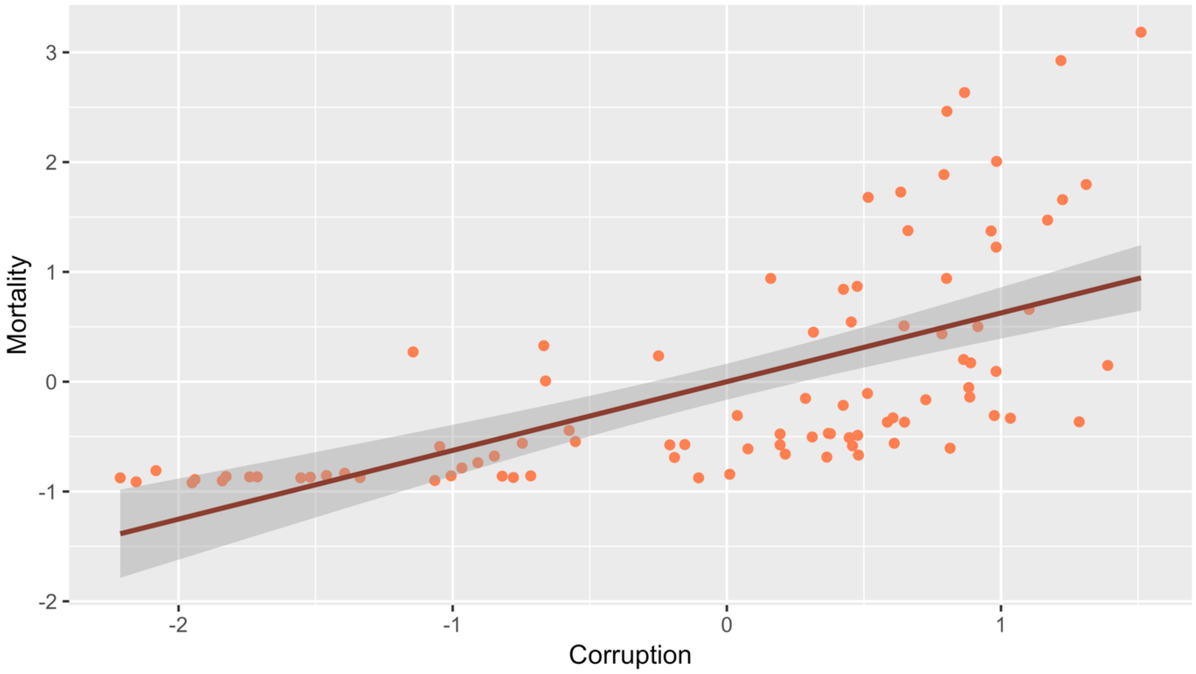

Figure 1: Scatter plot of mortality and corruption with linear regression line (dark red) and confidence interval (grey).

(c) i.

It is similar as we have measurement error in the dependent variable y. Considering the possible errors made by extracting the numbers of deaths it seems plausible that these would satisfy the classical measurement error condition. This would not lead to any bias in the estimate and we would expect an estimate in the same range as before. A problematic "input" to the measurement error might arise from the difference between the actual child mortality and the child mortality within hospitals. It seems quite likely that this error might depend on corruption, as one would expect less infrastructure in countries with high corruption and therefore a lower rate of women that give birth in hospitals. The variable is therefore not likely to satisfy the conditions of classical measurement error, as $Cov(\eta_i, x_i) \neq 0$. We expect the coefficient in a regression of corruptionun on hospital_deaths to be in the same range or even lower if the classical measurement error condition is not fulfilled.

ii.

```
===============================================
                        Dependent variable:
                    ---------------------------
                          hospital_deaths
-----------------------------------------------
corruptionun                 0.528***
                              (0.091)

Constant                      0.00000
                              (0.090)

-----------------------------------------------
Observations                    90
R2                             0.279
Adjusted R2                    0.271
Residual Std. Error       0.854 (df = 88)
F Statistic           34.057*** (df = 1; 88)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

The new estimate with 0.528 is lower than before (0.626) but still of the same magnitude. This comes at no great surprise, as we would expect some deviation but no bias. As we expected, there is a negative bias as our estimate in the above table is smaller than in our regression table in exercise 2.1.b.

iii.

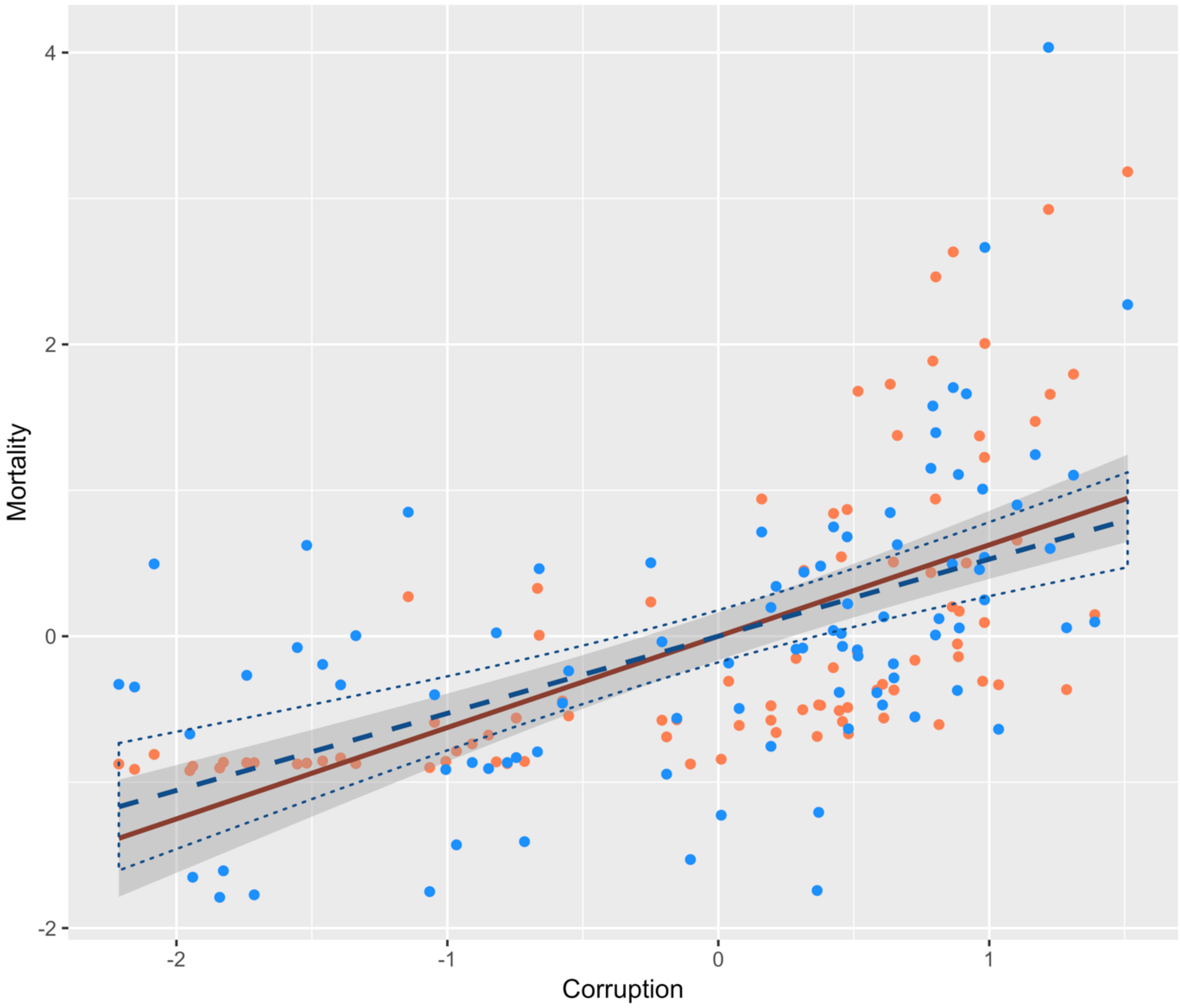

Figure 2: Scatter plot of mortality (UN: orange, Hospital deaths: blue) and corruption with population regression line (dark red) and its confidence interval (grey). Regression line from regression with hospital deaths in blue (dashed) as well as its confidence intervall (dotted).

The new regression line (blue, dashed) lies within the confidence interval of the population regression line (grey area). This supports the view that the measurement error behaves "nicely" and the estimate is unbiased. Also in line with the calculations in ex. 1.2 is the increased confidence interval for the regression line as the standard error for the estimates increases.

(d) i.

```
===============================================
                        Dependent variable:
                    ---------------------------
                            mortalityun
-----------------------------------------------
ruleoflaw                    0.361***
                              (0.099)

Constant                       0.000
                              (0.099)

-----------------------------------------------
Observations                    90
R2                             0.131
Adjusted R2                    0.121
Residual Std. Error       0.938 (df = 88)
F Statistic           13.215*** (df = 1; 88)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

We regressed *mortalityun* on *ruleoflaw* which serves as a proxy for corruption. We assume that this proxy has a measurement error and is different from the "true" *corruption* estimate. Our estimated coefficient for *ruleoflaw* is 0.361 which is indeed smaller than our estimates in the previous regressions. Furthermore, the standard error increased slightly. We conclude that this is due to the attenuation error that is negative when the estimated beta coefficient has a positive slope. The larger standard error is due to the covariance between the true of x_i and of η_i that has to be taken in to account.

(e) i.

It is similar to Question 2 as we have measurement error in the dependent variable y. The problem lies again in the difference between the actual child mortality and the child mortality reported by the governments. It seems quite likely that this error might depend on corruption, as one would expect countries with high corruption to be less precise about their true mortality rate and even systematically report lower figures. It is therefore very likely that the conditions of classical measurement error are not fulfilled.

ii.

Our Equation from Question 3b) sates the bias of the estimated $\hat{\beta}$ with as follows:

$$\hat{\beta} = \beta - \frac{Cov(\eta_i, x_i)}{Var(\epsilon_i)Var(x_i)}$$

As stated in i), one would expect the covariance between the measurement errors in η_i and x_i to be positive as the errors are expected to be larger for countries with a higher level of corruption.) We therefore expect a negative bias in the estimate.

iii.

```
===============================================
                        Dependent variable:
                    ---------------------------
                              govmort
-----------------------------------------------
corruptionun                  0.358***
                              (0.100)

Constant                      0.00000
                              (0.099)

-----------------------------------------------
Observations                    90
R2                             0.128
Adjusted R2                    0.118
Residual Std. Error       0.939 (df = 88)
F Statistic           12.902*** (df = 1; 88)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

As expected yields a regression with the self-reported mortality ($\hat{\beta}_{self} = 0.358$) a much lower estimate than the regression with the UN estimates ($\hat{\beta}_{UN}$= 0.626). This indicates the proposed negative bias stated in exercise ii).

(f)

We argue that our model in 1d) is the most dangerous in terms of identification of the true causal effect of corruption on child mortality because in this model we have measurement error in the independent variable x whereas in 1c) and 1e) we only have measurement error in the dependent variable y. In the third paper and pencil question we showed that a measurement error in y is only problematic when the the assumptions for *classical measurement errors do* not hold. In other word, we can expect that our estimated coefficient is unbiased when these classical measurement errors hold. On the other hand, when we have measurement error in the independent variable, the coefficient will be biased (attenuation bias) even if the assumption of *classical measurement errors* do hold. Therefore the model in 1d) is most likely to be biased even though our argumentation is theoretical and it is highly questionable that the classical measurement errors hold for model 1c) and 1e). Nevertheless, it would be very difficult to answer this question in reality as the “true relation” would remain unobservable which makes it difficult to assess the magnitude for each bias even if we would manage to collect data for our models in 1c) - 1e).

3 R Code

```
library(data.table)
library(stargazer)
library(ggplot2)
# load data
dataPS3= read.csv(file.choose(), header = T, sep = ",", dec = ".")


# Ex 2.1b.i regression and plot ------------
reg1 <- lm(mortalityun ~ corruptionun, data = dataPS3)
stargazer(reg1, type="text")
# claculate t stat and p value for the beta coef
t_stat =  (coef(summary(reg1))[2,1] - 0) / coef(summary(reg1))[2,2]
p_value = 1 - pt(t_stat, df = 88)


# Ex 2.1b.iii regression and plot ------------
chart_true_reg <- ggplot(dataPS3) +
  geom_point(aes(corruptionun, mortalityun),
             colour = "coral") +
  geom_smooth(mapping = aes(corruptionun, mortalityun),
              method = "lm", se=TRUE,
              colour="coral4") +
  labs(y = "Mortality", x = "Corruption")
print(chart_true_reg)
```

```
#save plot
ggsave("Ex_2_1biii.png",
       width = 7,
       height = 4,
       dpi = 600,
       path = paste0("/Users/Manu/Documents/Uni/Master/herbstsemester 2018/empirical methods/My PS 3"))


# Ex 2.1c.ii regression  and plot ------------
reg2 <- lm(hospital_deaths ~ corruptionun, data = dataPS3)
stargazer(reg2, type="text")


# Ex 2.1c.iii regression  and plot ------------
chart_1c <- ggplot(dataPS3) +
  geom_point(aes(corruptionun, mortalityun),
             colour = "coral") +
  geom_smooth(mapping = aes(corruptionun, mortalityun),
              method = "lm", se=TRUE,
              colour="coral4") +
  geom_point(aes(corruptionun, hospital_deaths),
             colour = "dodgerblue") +
  geom_smooth(mapping = aes(corruptionun, hospital_deaths),
              method = "lm",
              se=FALSE, colour = "dodgerblue4",
              linetype = "dashed", fill = "NA") +
  stat_smooth(mapping = aes(corruptionun, hospital_deaths),
              method="lm",
              fill=NA, colour="dodgerblue4",
              linetype=3, geom="ribbon") +
  labs(y = "Mortality", x = "Corruption")
# print plot
print(chart_1c)
#save plot
ggsave("Ex_2_1ciii.png",
       width = 7,
       height = 6,
       dpi = 600,
       path  =  paste0("/Users/Manu/Documents/Uni/Master/herbstsemester
2018/empirical methods/My PS 3"))

# Ex 2.1d.i regression  ------------
reg3 <- lm(mortalityun ~ ruleoflaw, data = dataPS3)
stargazer(reg3, type="text")

# Ex 2.1e.iii regression of govmort  ------------
reg4 <- lm(govmort ~ corruptionun, data = dataPS3)
stargazer(reg4, type="text")
```