Problem Set 2 Empirical Methods

Patrick Glettig (13-252-143), Carmine Ragone (14-995-021)

05.11.2018

Paper and Pencil Questions

1 Coefficient Interpretation

a)

$$consumption_i = \beta_1 + \beta_2 income_i + \epsilon_i$$

A 1-unit increase in income (1 USD) would effect in an increased consumption of 0.267 USD.

b)

$$consumption_i = \beta_1 + \beta_2 income_i + \beta_3 famsize + \epsilon_i$$

A 1-unit increase in income (1 USD) would effect in an increased consumption of 0.254 USD, holding family size constant. Every additional family member increases consumption by 625 USD.

c)

The coefficient on income is still 0.254, so the effect size did not change from the model in b). But we can now interpret it as: a 1-unit increase in income (1 USD) would effect in an increased consumption of 0.254 USD, controlling for family size and house owners.

d)

The coefficients change from model (1) to model (2) as part of the coefficient in *income* also reflected $famsize$, a intuitive explanation would be that people with larger families have to work (& hence earn) more to feed the family which biases the effect of income. The coefficients $\beta_2$andβ_3 barely change from model (2) to model (3), so the fact that someone is owning a house did not explain any part of the coefficients on $income$and$famsize$. As this makes model (3) inefficent, model (2) is the right one since it allows causal statements on the effect of income and family size on consumption.

Omitted Variable Bias

a)

$\hat{\alpha}_1$ is defined as:

$$\hat{\alpha}_1 = \frac{cov(X_{1i}, Y_i)}{var(X_{1i})}$$

Plug in Y_i of the true model to find the conditional expectation:

$$E(\hat{\alpha}_1|X) = \frac{cov(X_{1i}, \beta_0) + cov(X_{1i}, \beta_1 X_{1i}) + cov(X_{1i}, \beta_2 X_{2i}) + cov(X_{2i}, \epsilon_i}{var(X_{1i})}$$
$$= \beta_1 + \beta_2 \frac{cov(X_{1i}, X_{2i})}{var(X_{1i})} = \beta_1 + \beta_2 \beta_{X_{2i} \hat{o} n X_{1i}}$$

Empirical Part

```
indicators <- fread("indicators.csv")
head(indicators)  # Check the format
dim(indicators)  #Check length
summary(indicators)  #Explore the data
```

b)

```
Model1 = mortalityun ~ corruptionun
lm1 <- lm(Model1, data = indicators, x = TRUE)
```

Table 1: Regression Results UN

	Dependent variable:
	Mortality
Corruption	0.626^{***}
	(0.083)
Constant	0.00000
	(0.083)
Note:	*p<0.1; **p<0.05; ***p<0.01

i)

The p-value for a one-sided test is simply