Dingyuan Zhang

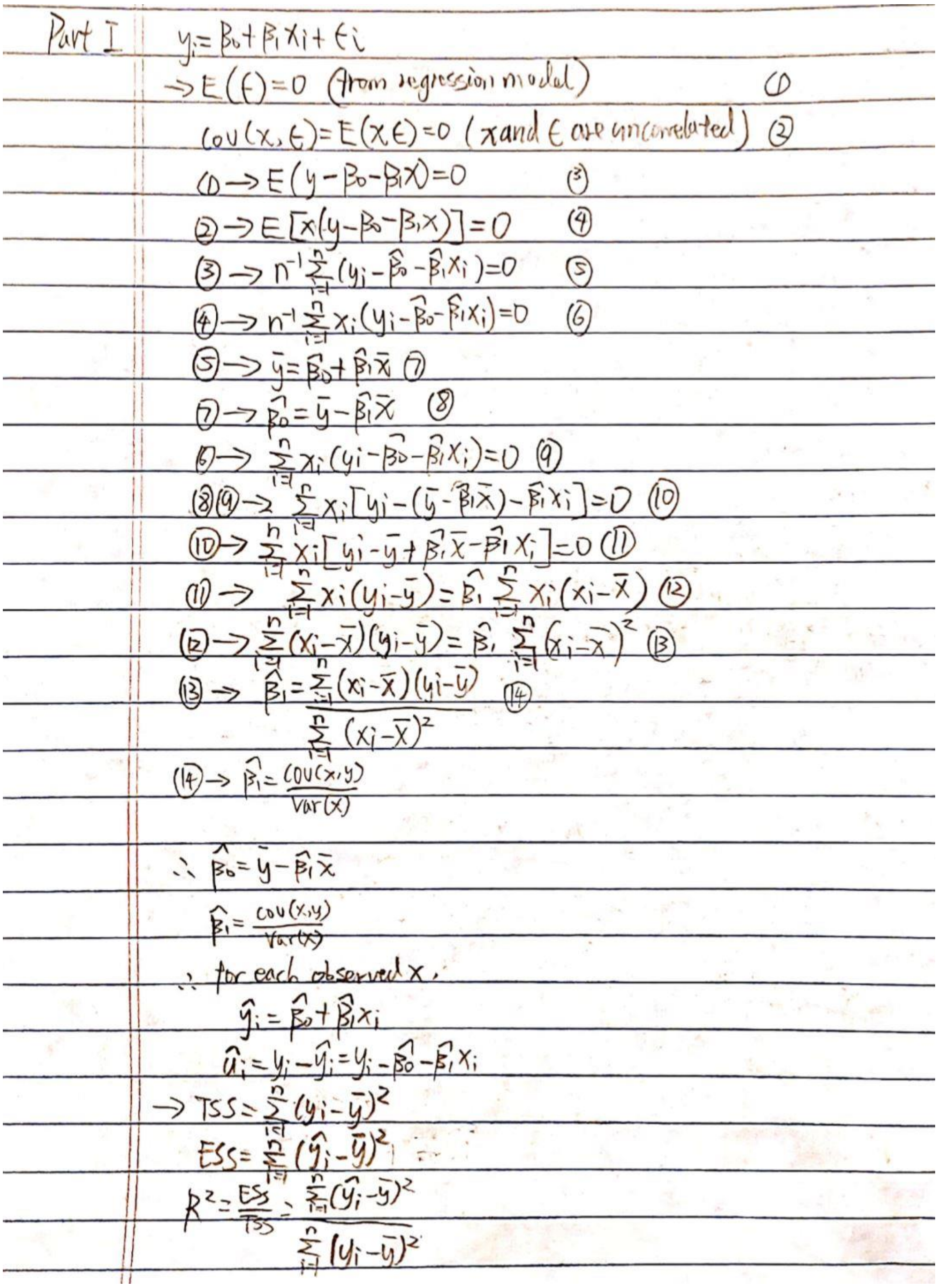
ECON320

Professor O'Connell

2019.9.25

Problem Set 2

Part I:



Part I $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$\rightarrow E(\epsilon) = 0$ (from regression model) ①

$cov(x, \epsilon) = E(x\epsilon) = 0$ (x and ϵ are uncorrelated) ②

① $\rightarrow E(y - \beta_0 - \beta_1 x) = 0$ ③

② $\rightarrow E[x(y - \beta_0 - \beta_1 x)] = 0$ ④

③ $\rightarrow n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ ⑤

④ $\rightarrow n^{-1} \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ ⑥

⑤ $\rightarrow \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ ⑦

⑦ $\rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ ⑧

⑥ $\rightarrow \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$ ⑨

⑧⑨ $\rightarrow \sum_{i=1}^{n} x_i [y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i] = 0$ ⑩

⑩ $\rightarrow \sum_{i=1}^{n} x_i [y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i] = 0$ ⑪

⑪ $\rightarrow \sum_{i=1}^{n} x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} x_i (x_i - \bar{x})$ ⑫

⑫ $\rightarrow \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^{n} (x_i - \bar{x})^2$ ⑬

⑬ $\rightarrow \hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$ ⑭

⑭ $\rightarrow \hat{\beta}_1 = \dfrac{cov(x,y)}{Var(x)}$

$\therefore \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\beta}_1 = \dfrac{cov(x,y)}{Var(x)}$

$\therefore$ for each observed x,

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

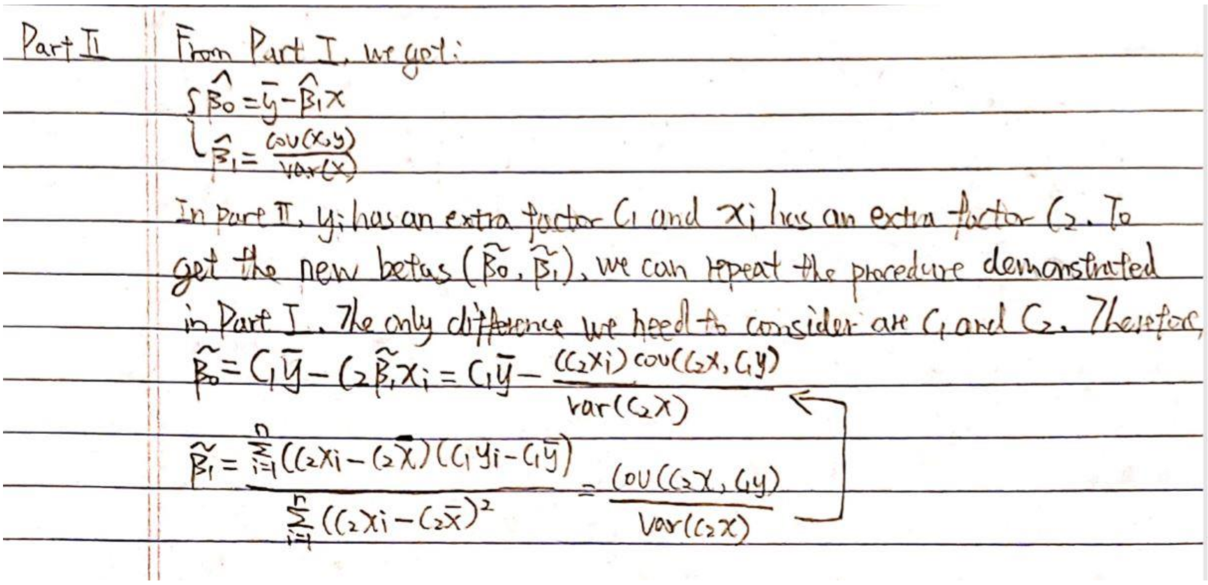
$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

$\rightarrow TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$

$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Part II:



Part II

From Part I, we got:

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x \\ \hat{\beta}_1 = \frac{cov(x,y)}{var(x)} \end{cases}$$

In part II, y_i has an extra factor C_1 and x_i has an extra factor C_2. To get the new betas $(\tilde{\beta}_0, \tilde{\beta}_1)$, we can repeat the procedure demonstrated in Part I. The only difference we need to consider are C_1 and C_2. Therefore,

$$\tilde{\beta}_0 = C_1\bar{y} - C_2\tilde{\beta}_1 x_i = C_1\bar{y} - \frac{(C_2x_i)\, cov(C_2x, C_1y)}{var(C_2x)}$$

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(C_2x_i - C_2\bar{x})(C_1y_i - C_1\bar{y})}{\sum_{i=1}^{n}(C_2x_i - C_2\bar{x})^2} = \frac{cov(C_2x, C_1y)}{var(C_2x)}$$

Part III:

(i) The coefficient on log(*dist*), which is 0.132, indicates that for every 1% increase in *dist*, there will be 0.132%$ increase in price. 0.132 is also the elasticity of *dist* with respect of price.

The sign of this estimate is what I expect it to be. The reason is due to the fact that incinerators produce significant amount of emission, which can negatively affect the living conditions of residents nearby. People prefers good living conditions, so they would choose to stay away from incinerators. Therefore, houses that are located further away from incinerators are very likely to have higher prices compare to those with similar conditions that are located near incinerators. This predication of mine is consistent with the model, in which increase in *dist* will result in increase in price.

(ii) I don’t think this simple regression provides an unbiased estimator of the ceteris paribus elasticity of price with respect to distance. The prices of houses can be affected by numerous factors, such as size, age, safety of nearby area, etc. However, the definition of ceteris paribus indicates that we need to hold other factors unchanged when we examine the effect of one factor. This can not be achieved with only one factor considered in this simple regression model.

(iii) For example:

1. Age of the house

   Under most circumstances, people prefer houses that are built more recently. Houses that are built 5 years ago are more likely to have better utilities and constructions than those that are built 100 years ago. Better utilities usually lead to higher living conditions, and less malfunctions. The age of the house is not correlated with distance from the incinerator.

2. Distance to good educational institutions

   Many people purchase new houses due to the education for their children or themselves. Houses near outstanding schools and prominent colleges are very likely to have higher prices than those that are far away from these institutions. The distance to good educational institutions can be related to the distance to incinerators of houses. It is common sense that incinerators are very likely to be located far away from schools and colleges, due to legal and ethical reasons.

Part IV:

(i) In this model, [beta hat 1] indicates that if the cigarettes consumption goes up by 1 unit, the expected birth weight of the baby will reduce by 0.514 unit

(ii) In this model, [beta hat 0] is the part of infant birth weight that is not affected by change in cigarettes consumption of the mother. This is not necessarily logical, because if a mother consumes too much cigarettes, miscarriage might happen and there won't be any data for infant birth weight.

(iii) Based on the regression, if a random mother smokes 10 cigarettes more per day than another mother, the infant birth weight of the former mother would be 10 * (0.514) = 5.14 unit less than the infant birth weight of the latter mother.

(iv) Based on this particular discovery, the residual for this particular situation can be calculated, which is 3 – 0.514 = -0.214 ounces. However, this might not be a accurate indication of the residual in general. This residual value is only calculated based on a segment of 10 cigarettes per day and two specific individuals. Therefore, the residual calculated in this scenario is not representative enough. If the segment becomes 5 cigarettes, the calculated result of residual can be drastically different.

(v) The ceteris paribus condition is not satisfied here. There are many different factors that are not taken into considerations, and some of the factors are related with the effect of cigarettes consumption. For example, if we take alcohol consumption into account, the overall health effect of cigarettes and alcohol might be different from simply adding up the effects of cigarette consumption and alcohol consumption. If this is the case, ceteris paribus will not hold. Therefore, the applied estimator here might not be unbiased.