Problem Set #1

MACSS 30100

Luxi Han, 10449918

Problem 1

(a) I'm going to choose *Agricultural Productivity and Structural Transformation, Evidence from Brazil* authored by Paula Bustos, Bruno Caprettini and Jacopo Ponticelli, from *American Economic Review* 2016 June Issue.

(b) The full citation from the website of *American Economic Review* is

Bustos, Paula, Bruno Caprettini and Jacopo Ponticelli. 2016. "Agricultural Productivity and Structural Transformation: Evidence from Brazil." *American Economic Review*, 106(6): 1320-65.

(c) The paper tries to understand how different factor biased technology changes affect the productivity and employment share in both agricultural and industrial sectors. The theoretical model is as follow:

$$Q_m = A_m L_m \tag{0.1}$$

$$Q_a = A_a[\gamma(A_L, L_a)^{\sigma-1/\sigma} + (1-\gamma)(A_T T_a)^{[\sigma-1]/\sigma}]^{\sigma-1/\sigma} \tag{0.2}$$

Where subscript L represents labor augmenting technology change and subscript T represent land augmenting technology change.

The empirical statistical model considered in this article is the following model:

$$\Delta y_j = \Delta\alpha + \beta\Delta A_j^{soy} + \gamma\Delta A_j^{maize} + \sigma Rural_{j,1991} + \theta X_{j,1991} + \Delta\epsilon_j \tag{0.3}$$

In the model, the subscript j represents the j_{th} muniplicity. The data takes the first difference form. The outcome of the model is denoted by Δy_j, which is the change in output per worker and change in employment share in either agricultural and industrial sector. ΔA_j^{soy} represents the increase in labor productivity of soybean production if adopting the labor augmenting GE soy technology; ΔA_j^{maize} represents the increase in labor productivity of maize production if adopting the land augmenting second season harvesting technology in maize production. $Rural_{j,1991}$ represents the rural population share in the $j_t h$ muniplicity in 1991 and $X_{j,1991}$ represents other muniplicity characteristics as controls.

(d) In this model, the exogenous variable are the parameters in front of the variables; additionally, in the model the labor and land endowment and elasticity of substitution in agricultural sector are considered exogenous variables in the model (ie. output variables is not affected by factor inputs).

Endogenous variables are: output per worker and employment share of agricultural and industrial sectors (outcome variables), the productivity increase in adopting GE soybean technology and second season harvesting technology in maize production. Rural population share and other municipal characteristics are also endogenous variables though they are controls.

(e) This model is a static, linear and stochastic model.

(f) Another variable i think would be biasing the result is the change of size of the down-stream industries that require soybean or maize as inputs. Take soybean as an example, the increase in the downstream industry (like feedstuff production) can result in a up-ward bias in the effect of ΔA_j^{soy} on output and may lead to a downward bias in the effect of ΔA_j^{soy} on employment share change.

Problem 2

(a) I model the life expectancy of a musician from four perspectives: genetic reason, lifestyle, family and career:

$$log(life\,expectancy) = \underbrace{\beta_{1,1}parents\,disease + \beta_{1,2}Gender + \beta_{1,3}Race+}_{\text{genetic reason}}$$
$$\underbrace{\beta_{2,1}log(num\,drug\,related\,news) + \beta_{2,2}log(weight/height) + \beta_{2,3}log(num\,year\,education)}_{\text{lifestyle}}$$
$$\underbrace{\beta_{3,1}married + \beta_{3,2}log(num\,children)+}_{\text{family}}$$
$$\underbrace{\beta_{4,1}log(income) + \beta_{4,2}log(income)^2 + \beta_{4,3}log(num\,concert\,/year) + \beta_{4,4}genre}_{\text{career}}$$

$$(0.4)$$

(b)   None

(c)   None

(d) The key factors in this model are whether parents have a chronicle disease, gender, years of education, weight and height ratio, income and genre. The reason why theses variables are important is that: 1) parents' disease is a good indicator of genetic disease that is heritable; 2) years of education is a good proxy for self-discipline and health related knowledge; 3) weight and height ratio is a good proxy for exercise; 4) genre is a good indicator for stress related issues and different lifestyle.

(e) There are several reasons why I choose these variables. The first reason is that they are comprehensive. A person's life expectancy should be determined by genetic?innate and other postnatal reasons. Postnatal reasons can involve their own lifestyle, family environment and work environment. In this sense, they are comprehensive.

The second reason is that they are all very retrievable. All(except maybe parents' chron-icle diseases) variables can be found on internet doing standard search and scraping.

The third reason is that they are good proxies. For example, it's hard to get information about work related pressure. Then the three career related variables may be good proxies. Income can correspond positively to pressure. But taking into account of the positive effect of career achievement on life expectancy, the effect of income may exhibit

a bell curve shape. Number of concerts per year may also be a good proxy for workload. Different artists of different genres can have totally different lifestyle. It's blatantly true that drug culture is prevalent for a certain group of artists in a certain genre. Additionally, lifestyle is something that is hard to get information on. Then years of education can turn out to be a good proxy for their degree of self-discipline and their awareness of the positive effect of a healthy lifestyle.

(f) The way I would test my hypothesis is that I can scrape the information particularly the variables that I propose in the linear regression equation. A first test would be to see the correlation between life expectancy and each variable. We should pay attention to whether the sign is the same as we hypothesize and how large the correlation is. Then we can train the model. I would divide my sample into two groups: one training set and the other is the test set. I can run the regression and firstly use R^2 as a preliminary test. Then taking the parameters from the training set to test the fit of the model in the test set.