

CECN702

# Project

Saumitra Mazumder

500720916

Ryerson University

September 16, 2020

I hereby declare that I am the sole author of this work.

# 1 Introduction

We estimated the relationship between the change in employment status over a year for a worker and worker's age, sex and race from April 2006 to April 2007. The observations consisted of workers who reported as employed on the April 2006 "Current Population Survey". The workers were subsequently surveyed about their employment status a year later, April 2007.

The employment status of a worker ( $EMPLOYMENT_i$ ) takes the value of 1 for a worker who was employed during both survey times of April 2006 and April 2007 and 0 when unemployed in April 2007. Of the prescribed dependent regressors, sex ( $SEX_i$ ) and race ( $BLACK_i$ ) were binary. The former takes values of 1 if a worker identifies as female and 0 otherwise. The latter regressor takes values of 1 when the worker identifies as black and 0 otherwise. The worker's age ( $AGE_i$ ) is a continuous regressor. The data set contains 5220 observations.

We hypothesized that a race identification of black should not be associated with a difference in employment status when holding age and sex constant. Since  $EMPLOYMENT_i$  is binary the regular OLS regression could yield predictions greater than 1 or less than 0. We estimated our desired relationship using a non-linear probability logit model, thereby restricting our predictions to between 0 and 1.

Our model predicted that identifying as black decreases the probability that a worker will be employed, when all other regressors are kept equal.

## 2 Data and Model

The observations used for the model were taken from the "Data for Empirical Exercises" Student Resources companion for Stock and Watson's Introduction to Econometrics, Third Edition. The dataset contained observations from 5220 workers who identified as employed in the April 2006 "Current Population Survey" and their employment status one year later.

The dataset also contained addition regressors surveyed in April 2006, namely: age, female, married, race, union status, separate binary regressors for state locations (northeast, southwest, etc), and separate binary regressors for level of education (high school, college, etc), as well as weekly earnings.

Of interest to us in this survey was the relationship between the change in the employment status of a worker and if the worker identified as black when age and sex kept constant. As such, we discarded any data not required in our regression and only kept data regarding race, age, sex and of course the dependent regressor, employment. Furthermore, since we were interested in workers who identified as black, any identification for race that was not black in the original dataset was regarded as 0 out our dataset. These changes did not change the size of our dataset.

We set up an indicator function:

$$\overline{EMPLOYMENT} = \beta_0 + \beta_1 AGE + \beta_2 SEX + \beta_3 BLACK + u \quad (1)$$

Here  $(??)$ , is considered an underlying propensity for the *EMPLOYMENT* regressor to take the value of 1. We used the logistic distribution to find the probabilities. That is, we estimated a model of the form:

$$\begin{aligned} \Pr(EMPLOYMENT_i = 1 | AGE_i, SEX_i, BLACK_i) \\ = \Lambda(\beta_0 + \beta_1 AGE_i + \beta_2 SEX_i + \beta_3 BLACK_i) \end{aligned} \quad (2)$$

$$\text{Where } \Lambda(z) = \frac{1}{1 + e^{-z}}$$

Our method to estimate this model was to choose  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$  such that the log-

likelihood function (??) is maximized. The estimation was solved using statistical package R.

$$\ln L = \sum_{i=1}^{5220} \ln[\Pr(EMPLOYMENT_i = y_i | AGE_i, SEX_i, BLACK_i)] \quad (3)$$

In general, the coefficients of a logit regression cannot be interpreted in the same way they are in an OLS regression. We were interested in the marginal effect that an identification of black would have on employment; i.e., how much the conditional probability of the employment value changes when the value of race changes when we holding all other regressors constant.

To find this marginal effect, we noted that the race regressor is and binary set up the following discrete change function:

$$\begin{aligned} \Delta = & \Lambda \left( \hat{\beta}_0 + \hat{\beta}_1 AGE_i + \hat{\beta}_2 SEX_i + \hat{\beta}_3 BLACK_i \middle| BLACK_i = 1 \right) \\ & - \Lambda \left( \hat{\beta}_0 + \hat{\beta}_1 AGE_i + \hat{\beta}_2 SEX_i + \hat{\beta}_3 BLACK_i \middle| BLACK_i = 0 \right) \end{aligned} \quad (4)$$

Finally, we tested the hypothesis in the usual way,

$$\begin{aligned} H_0 : & \beta_3 = 0 \\ T = & \frac{\hat{\beta}_3 - \beta_3}{SE(\hat{\beta}_3)} \end{aligned} \quad (5)$$

at a significance of 5%

### 3 Results

We can see from our results in Table ??, that our estimated logit model is:

$$\begin{aligned}\widehat{\Pr}(EMPLOYMENT_i = 1|AGE_i, SEX_i, BLACK_i) \\ = \Lambda(1.798 + 0.015AGE_i - 0.441SEX_i - 0.186BLACK_i)\end{aligned}\quad (6)$$

Furthermore, our estimated marginal effects function is

$$\begin{aligned}\widehat{\Delta} = & \Lambda(1.798 + 0.015AGE_i - 0.441SEX_i - 0.186) \\ & - \Lambda(1.798 + 0.015AGE_i - 0.441SEX_i)\end{aligned}\quad (7)$$

Table 1: The Maximum likelihood estimates of age, sex and black and constant as computed by R as well as their respective standard errors.

<i>Dependent variable:</i>	
	employed
Age	0.015*** (0.004)
Sex	-0.441*** (0.093)
Black	-0.186 (0.161)
Constant	1.798*** (0.176)
Observations	5,220
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

We noticed that the p-value for the coefficient of the Black regressor was greater than 0.05, hence we could not reject  $H_0$  from (??).

Our model states that a black female worker with a mean age of 42 will have a probability

of being employed 0.8582715, where as a non black female worker with the same mean age will have a probability of being employed 0.8794511. Using our estimated marginal effects function (??), we find that the marginal effect is  $-0.1862355$  for a female worker. Our model predicts that a black male worker with a mean age will be employed with a probability of 0.9039293, whereas a non black, male worker with the same mean age will be employed with a probability of 0.9189304. Here, the marginal effect is  $-0.1862355$  for a male worker. Hence, we conclude that there is an effect of identifying as black on the probability of employment when holding other variables constant.