Team Project 1

Purple Team, presenting on Part II

Raza Lamb, Robert Wan, Erika Fox, Minjung Lee, Preet Khowaja

Introduction

This is an analysis to infer the relationship between job training for disadvantaged workers and their wages, from an experiment conducted at the National Supported Work (NSW) Demonstration.

Question of Interest:

Is there evidence that workers who receive job training tend to be more likely to have positive (non-zero) wages than workers who do not receive job training?

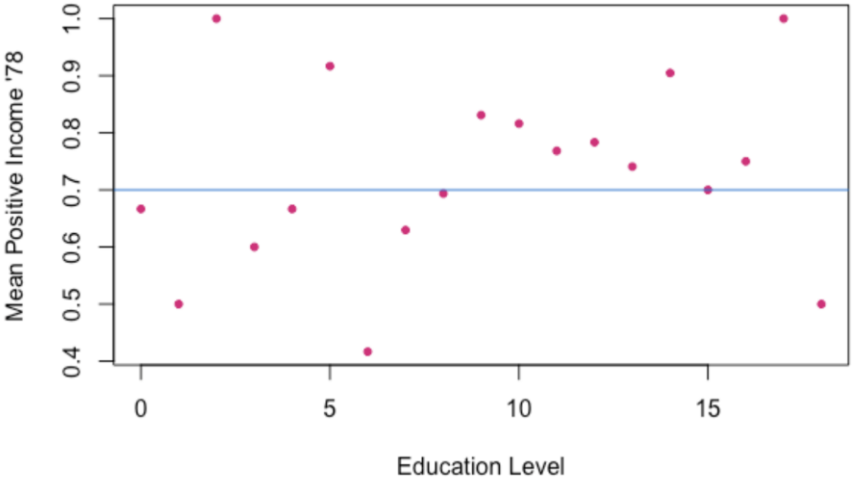
More specifically:

- Quantify the effect of the treatment, that is, receiving job training, on the odds of having non-zero wages.
- What is a likely range for the effect of training?
- Is there any evidence that the effects differ by demographic groups?
- Are there other interesting associations with positive wages that are worth mentioning?

Data

We created additional factor variables based on insights from the EDA:

- ***positive***: 1 if the participant had a positive (non-zero) income in 1978, 0 otherwise. **(the response variable)**
- ***zero***: 1 if the participant had a non-positive income (income of 0) in 1974, 0 otherwise.
- ***newed***: 1 if educ is greater than or equal to 9 years of education, 0 otherwise.



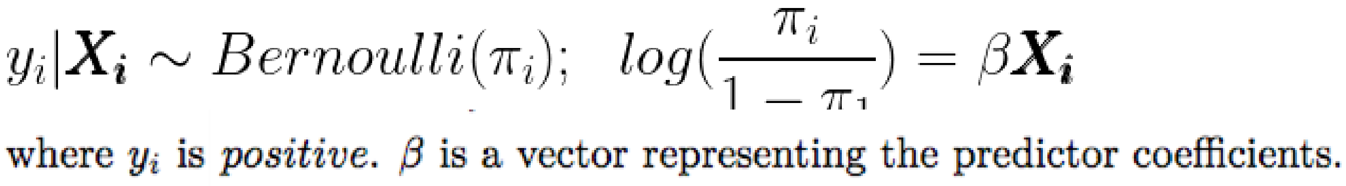
We decided to use *re74* as the baseline income variable. We did not use the variable *re75*. While the control group was selected based on income in 1975, the income for the treatment group is not comparable as some people began their training in 1975.

Data

“If you have any pretty pictures..”

Model, part 1: Selection

Selection methods: aic_backwards, aic_stepwise, and bic_stepwise

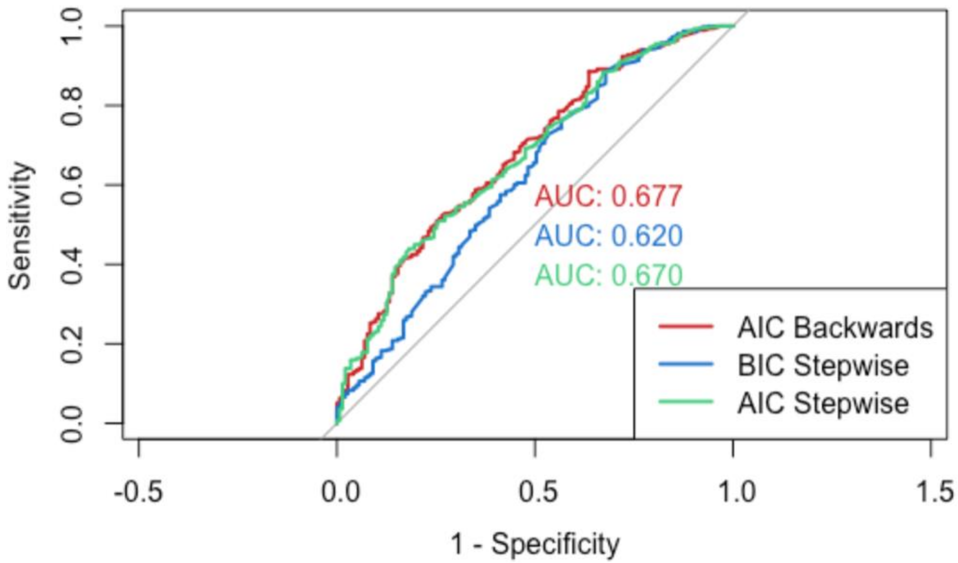


$$y_i|\boldsymbol{X_i} \sim Bernoulli(\pi_i); \quad log(\frac{\pi_i}{1-\pi_1}) = \beta\boldsymbol{X_i}$$

where y_i is *positive*. β is a vector representing the predictor coefficients.

- **Null Model predictors:** *treat*
- **Full Model predictors:** *treat:agec, treat:educc, treat:black, treat:hispan, treat:married, treat:re74c, treat:zero, treat:newed, black:re74c , re74c:married, educc:black, and educc:married*

Selection Results



AIC_Backwards vs AIC_Stepwise

- **AIC_Backwards**: the interaction of *treat:zero* is included and significant
- **AIC_Stepwise**: *treat* is significant in AIC_Stepwise, while it is not in AIC_Backwards

- We used Chi-squared tests to determine which model to use because the ROC curves are similar
- The test for BIC_Stepwise and AIC_Backwards revealed that the difference between them is significant enough for us to use AIC
- The difference between AIC_Backwards and AIC_Stepwise was not significant

AIC backwards results:

$$y_i|x_i \sim Bernoulli(\pi_i)\log(\frac{\pi_i}{1-\pi_i}) = x_i\beta,$$

where y_i is positive. x_i includes the predictors variables: *treat, agec, educc, black, re74c, zero, hispanic,* and *newed*, and the interactions *treat:agec, treat:hispanic,* and *treat:zero*. $\boldsymbol{\beta}$ is a vector representing the predictor coefficients.

However, during model assessment, we found a trend. So we added some transformations for our final model. We also removed two terms.

Final model:

same as model above with added *agec*^{2} and *agec*^{3} terms and removed *hispanic* and *treat:hispanic* terms

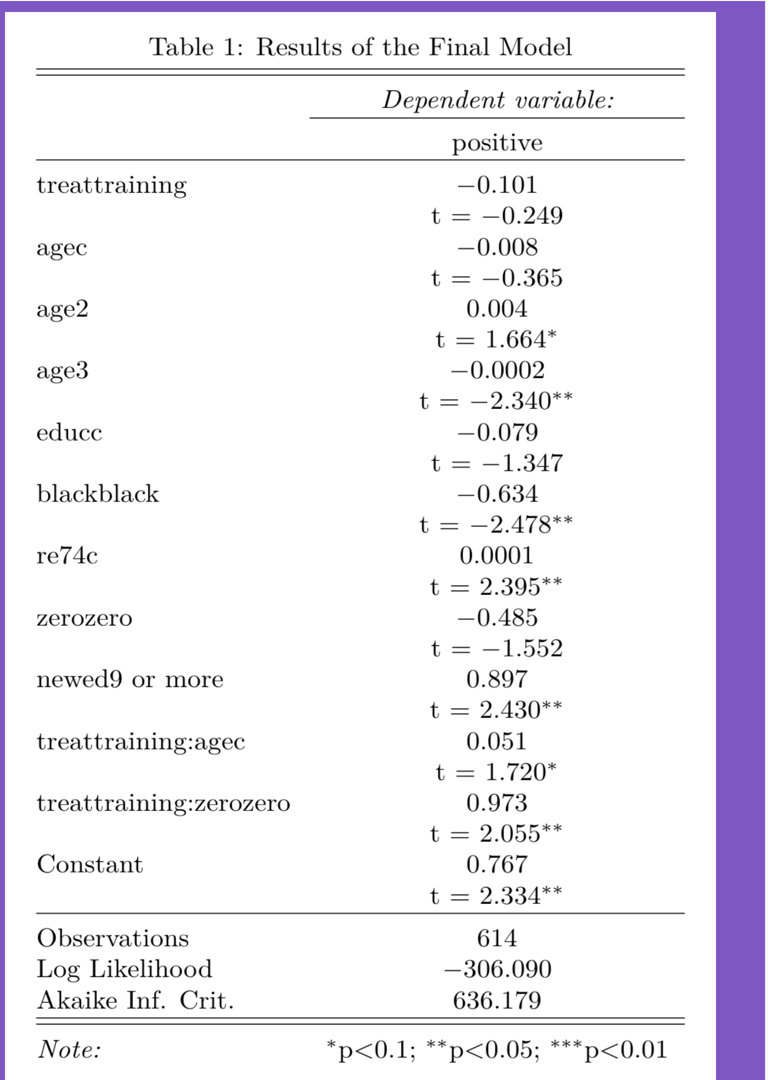
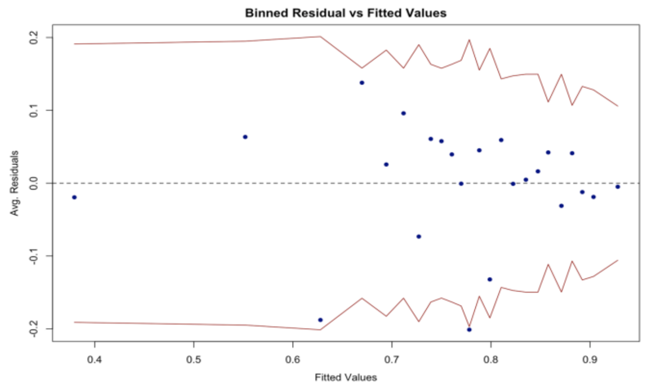


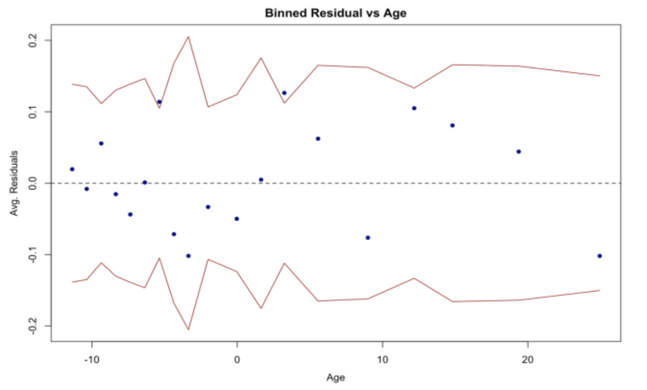
Table 1: Results of the Final Model

	Dependent variable:
	positive
treattraining	−0.101 t = −0.249
agec	−0.008 t = −0.365
age2	0.004 t = 1.664*
age3	−0.0002 t = −2.340**
educc	−0.079 t = −1.347
blackblack	−0.634 t = −2.478**
re74c	0.0001 t = 2.395**
zerozero	−0.485 t = −1.552
newed9 or more	0.897 t = 2.430**
treattraining:agec	0.051 t = 1.720*
treattraining:zerozero	0.973 t = 2.055**
Constant	0.767 t = 2.334**
Observations	614
Log Likelihood	−306.090
Akaike Inf. Crit.	636.179
Note:	*p<0.1; **p<0.05; ***p<0.01

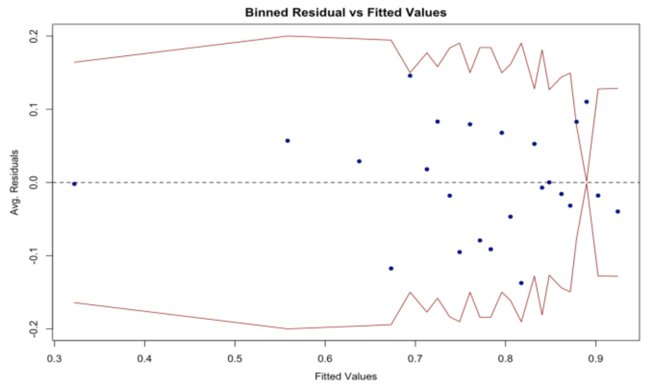
Model, part 2: Assessment

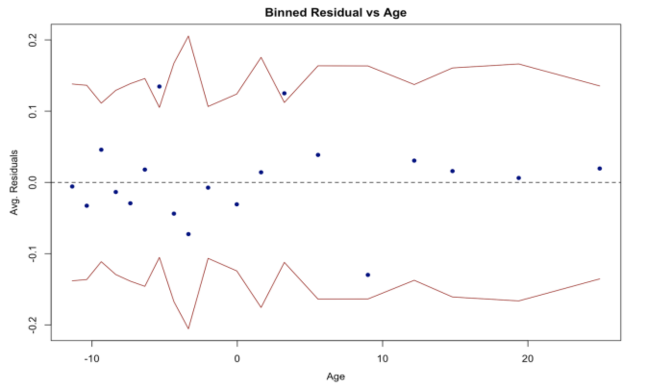
Before transformation:



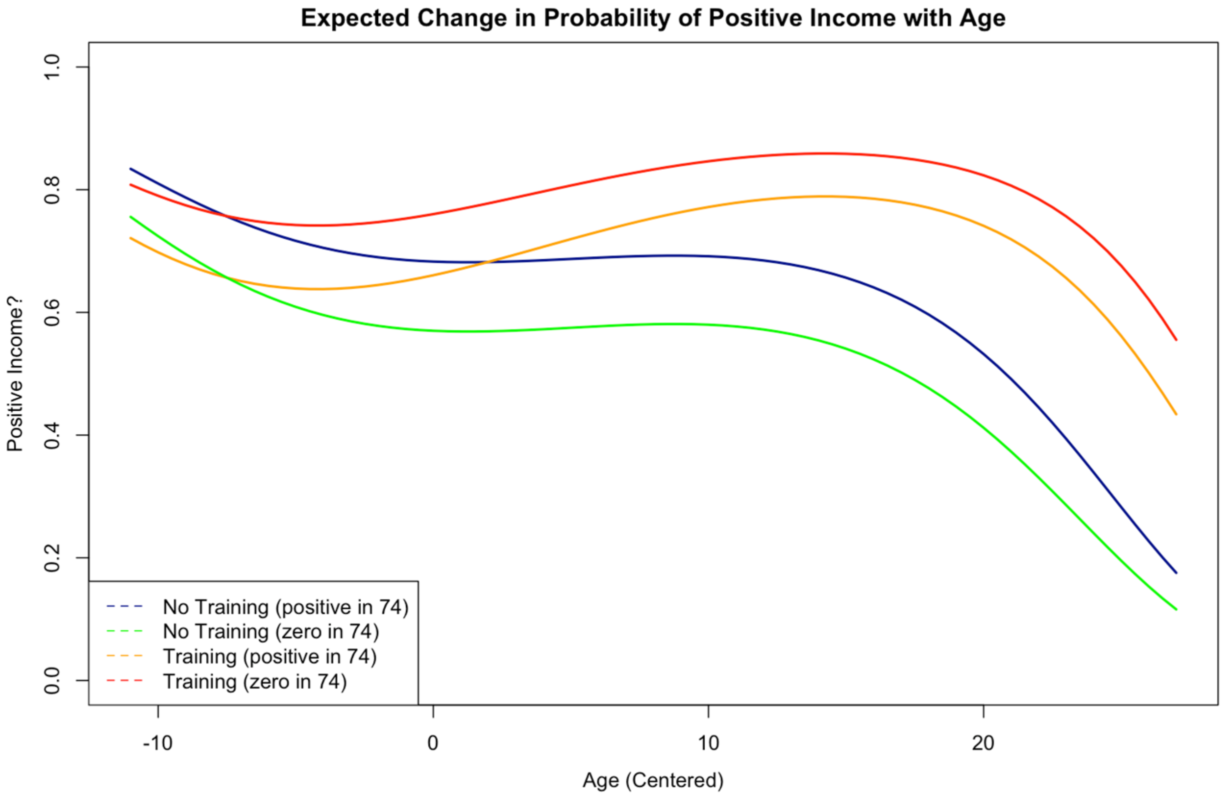


After transformation:





Interpretation and Conclusions

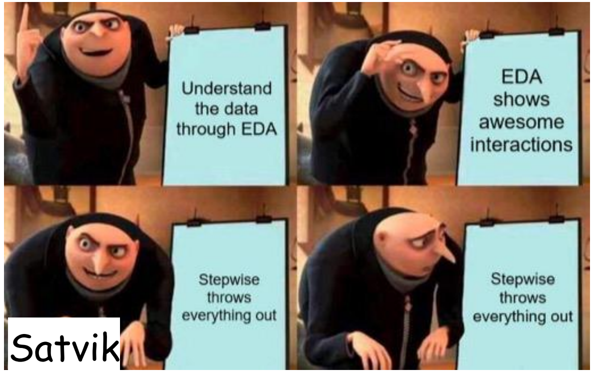


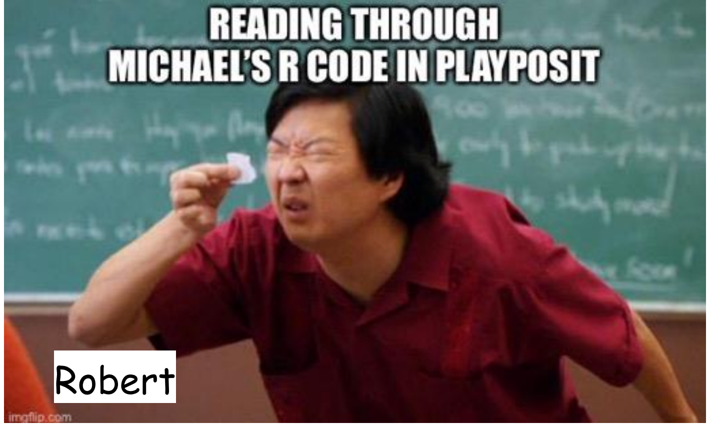
Limitations:

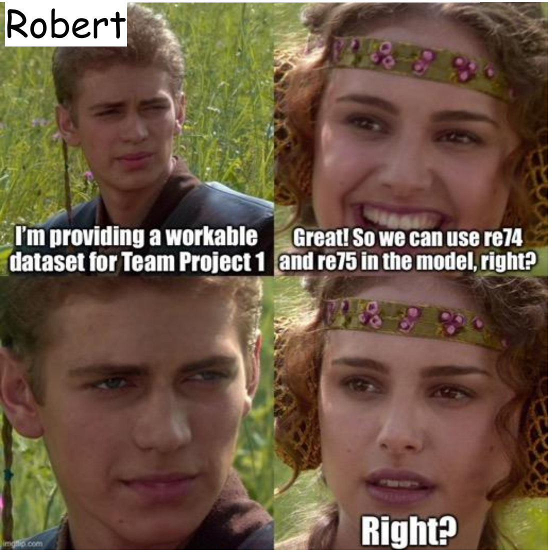
- Unable to use *re75* variable in our analysis, because of noise
- This interpretation is specific to the training program represented in this data.
- The control group might not have the same characteristics as the test group, because we selected them using different methods.
- Some categories were lacking in data (i.e. hispanic), prompting us to exclude the variable from our model.
- Modern inference about job training from this analysis is inappropriate as this data is from the 70’s, only includes men, etc.

A gallery of 702 memes

Moritz


Satvik
Understand the data through EDA
EDA shows awesome interactions
Stepwise throws everything out
Stepwise throws everything out


Robert


Robert
I'm providing a workable dataset for Team Project 1
Great! So we can use re74 and re75 in the model, right?
Right?

Raza

Preet