Mapping the Determinants of Trade Union Membership Using Explainable Artificial Intelligence Framework

Research Topic

- Main goal of this research is to map the individual level determinants of trade union membership by using Explainable Artifical Intelligence Framework.

Research Strategy

This Research Follows the Steps Below:

1. Variable Selection
2. Picking the Best Performing Model
3. Shapley Value Decomposition of the Model
4. Exploring the Relationships and Forming Hypotheses
5. Hypothesis Testing

Methodological Remarks

- Strictly Deductive approaches are only useful when a researcher has well defined theoretical models.
- It requires statistical skills to translate the theoretical model to the realm of statistics.
- Poorly defined statistical models provide misleading results.
- We might end up reproducing common-sense results!

Methodological Remarks

- Overfitting causes the problem of capturing the noise as part of the signal.
- Underfitting leads missing the signal in the data.
- Collinearity might flip the effect sizes.

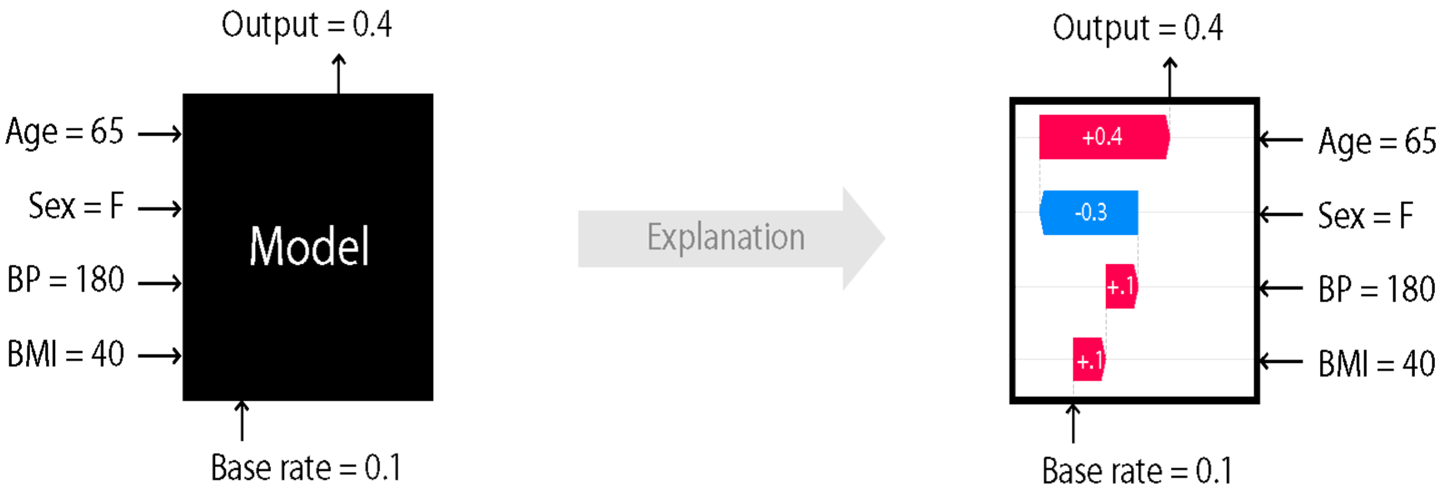
What are the Other Possibilities?

- Blackbox Models: Decision Trees, Random Forest, Support Vector Machines, Neural Networks, Gradient Boosting Machines etc.
- Blackbox models outperform linear models, but they are not interpretable!

LightGBM

- To pick the best performing model, Lightgbm, Catboost and xgboost algorithms were applied to the data in comparison with the Logistic Regression Model. All of them outperformed the logistic regression and gave similar results.

SHAP



How Does Shap Work?

- SHAP uses Shapley Values from game theory to break down variable contributions per each prediction that the model makes. Following 3 axioms bring fair credit allocation
- Additivity
- Null Player
- Monotonicity

1) Variable Selection

cntry	polintr	cptppola	rlgdgr	gndr	agea	chldhhe	eduyrs	estsz	tporgwk	hinctnta	hincfel	iincsrc	frprtpl	ifrjob	iprspot

Cntry: Country (18)

Polintr: Political Interest (4)

Cptppola: Confident in own ability to participate in politics(5)

Rlgdgr: How religious are you(11)

Gndr, agea,eduyrs

Chldhhe: Children living at home or not(2)

Estsz: Establishment size(5)

Tporgwk: What type of organisation work/worked for(6)

Hincfel: Feeling about household's income nowadays

Iincsrc: Source of Income(9)

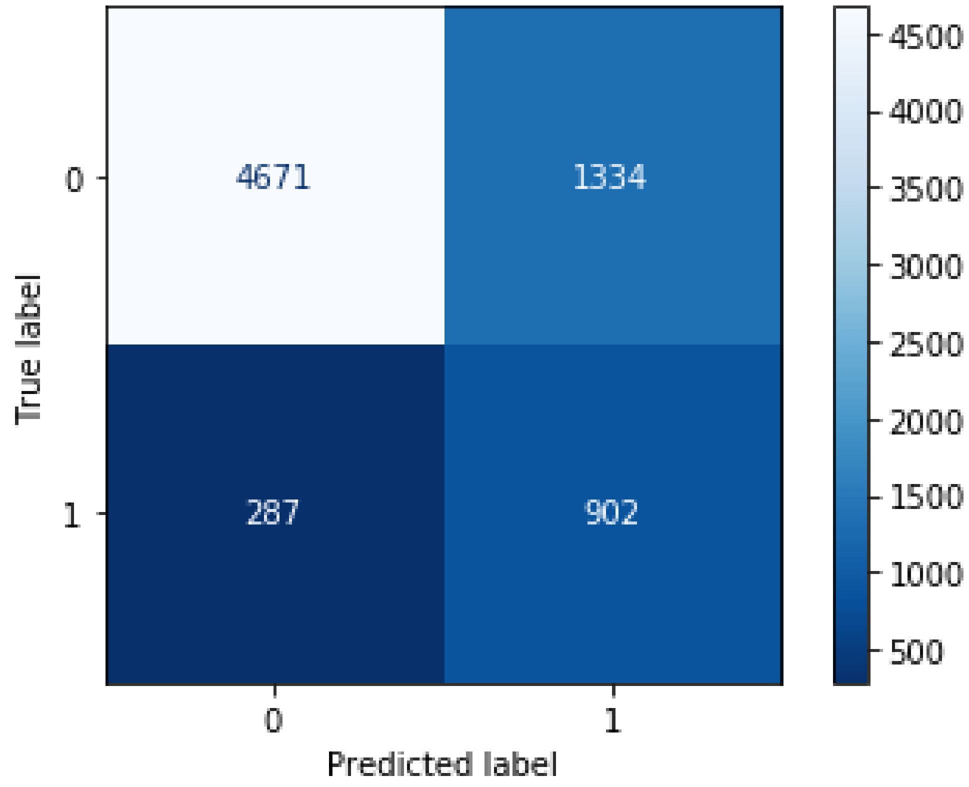
Frprtpl: Political system in country ensures everyone fair chance to participate in politics(5)

Ifrjob: Compared other people in country, fair chance get job I seek(11)

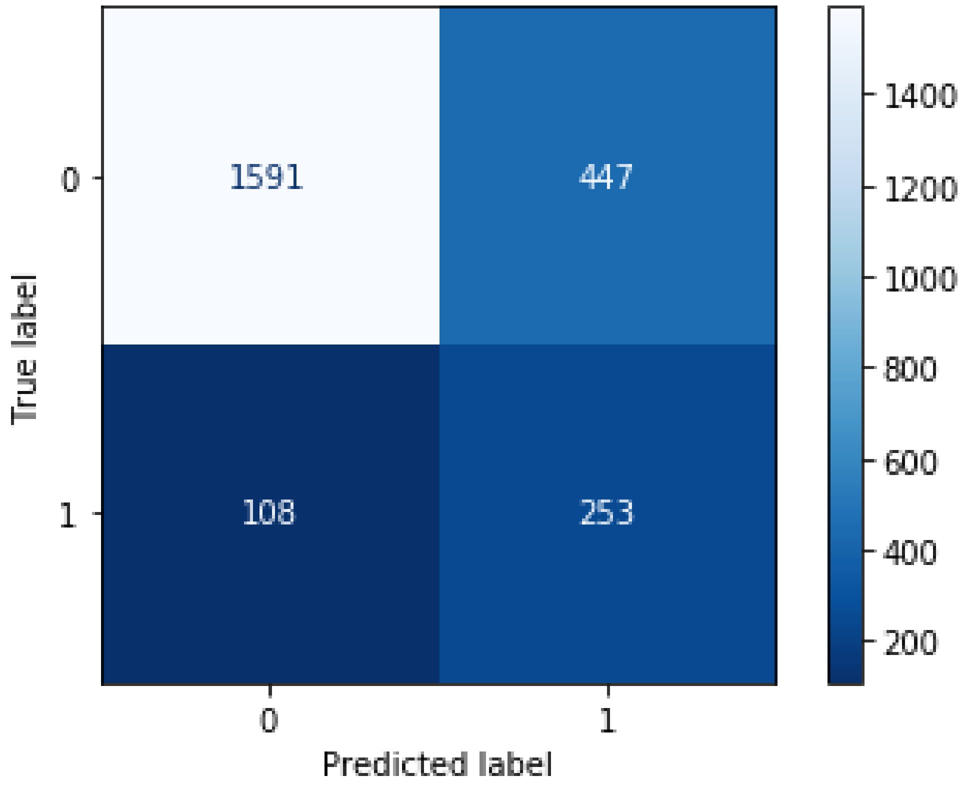
Iprspot: Important to get respect from others(6)

2)PICKING THE BEST PERFORMING MODEL

Lightgbm on Training Data

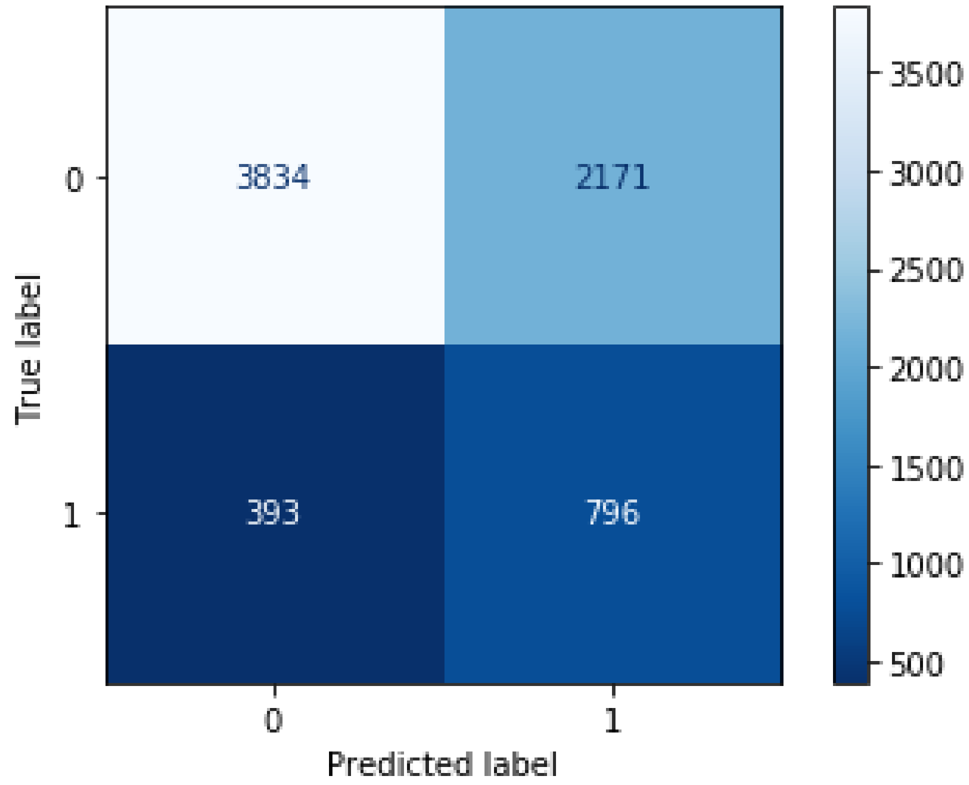


Lightgbm on Test Data

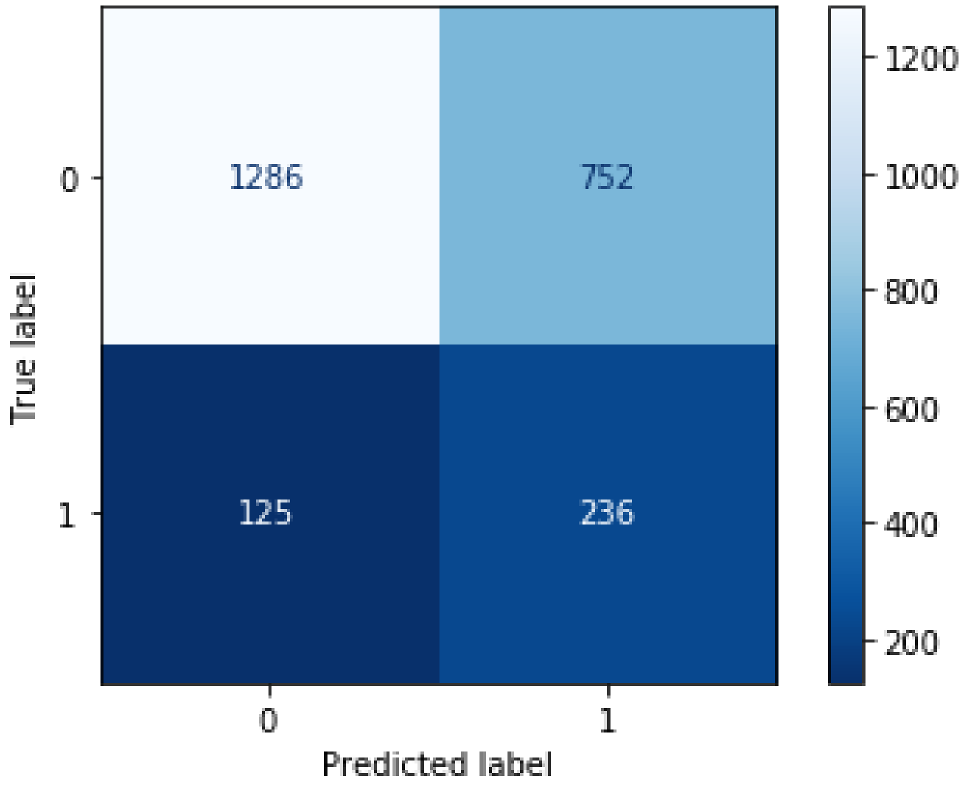


2)PICKING THE BEST PERFORMING MODEL

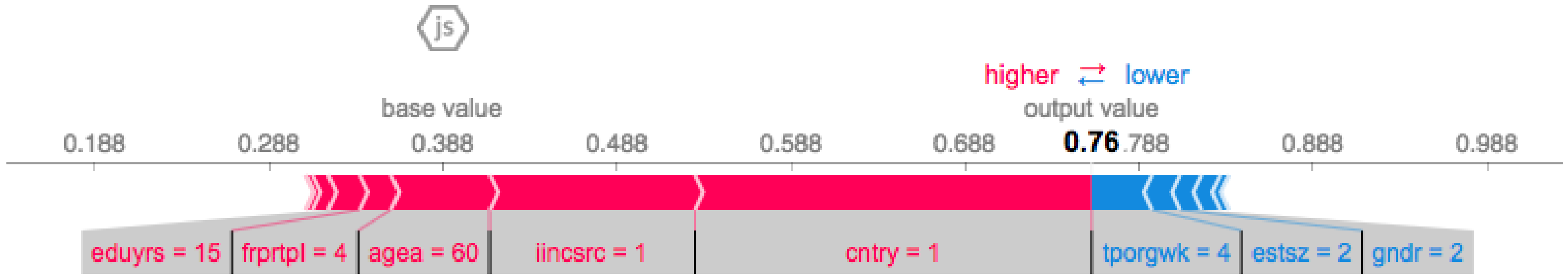
Logistic Regression on Training Dataset

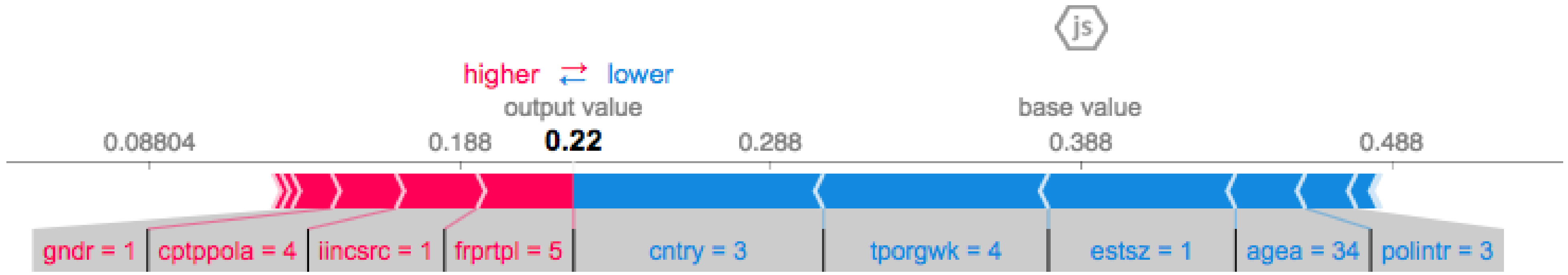


Logistic Regression Test Dataset

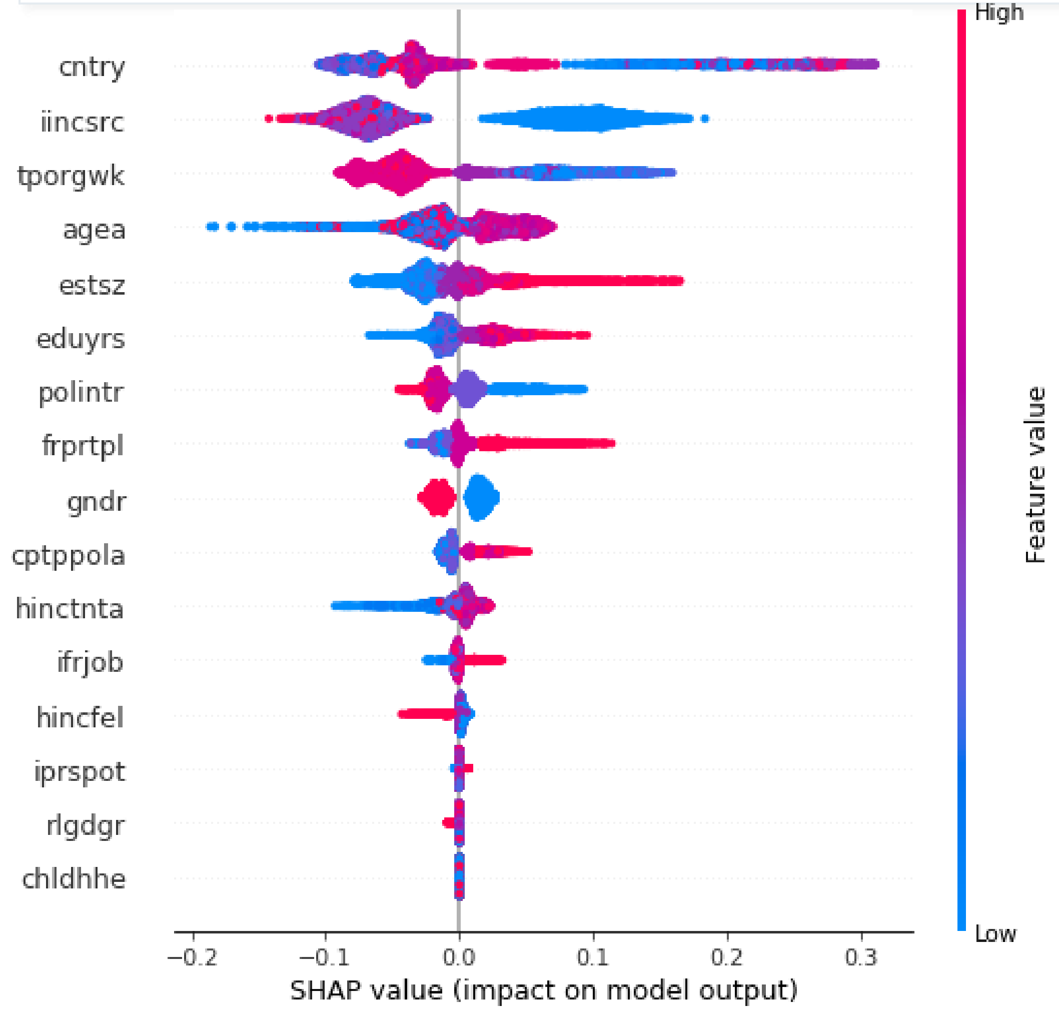


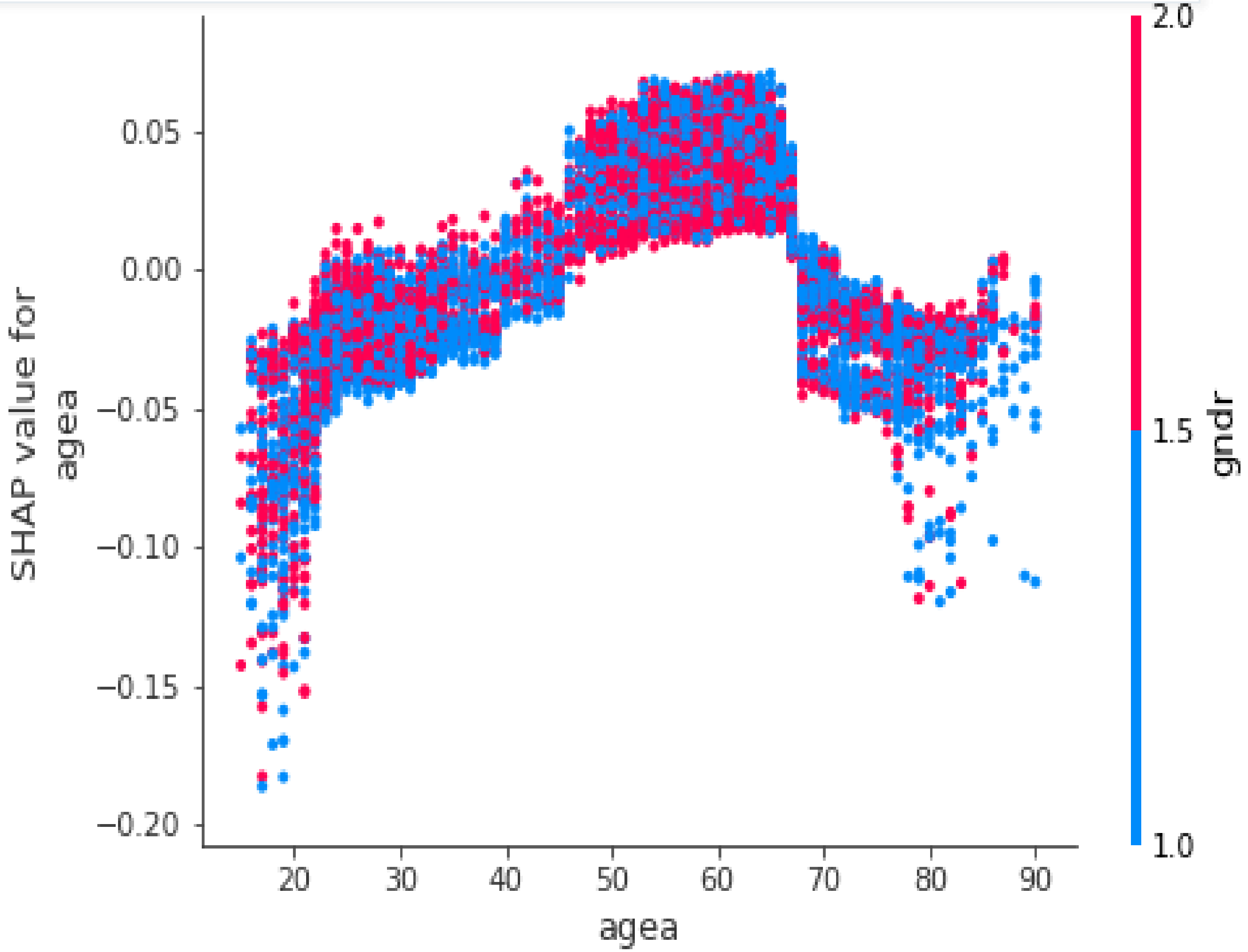
3) SHAPLEY VALUE DECOMPOSITION



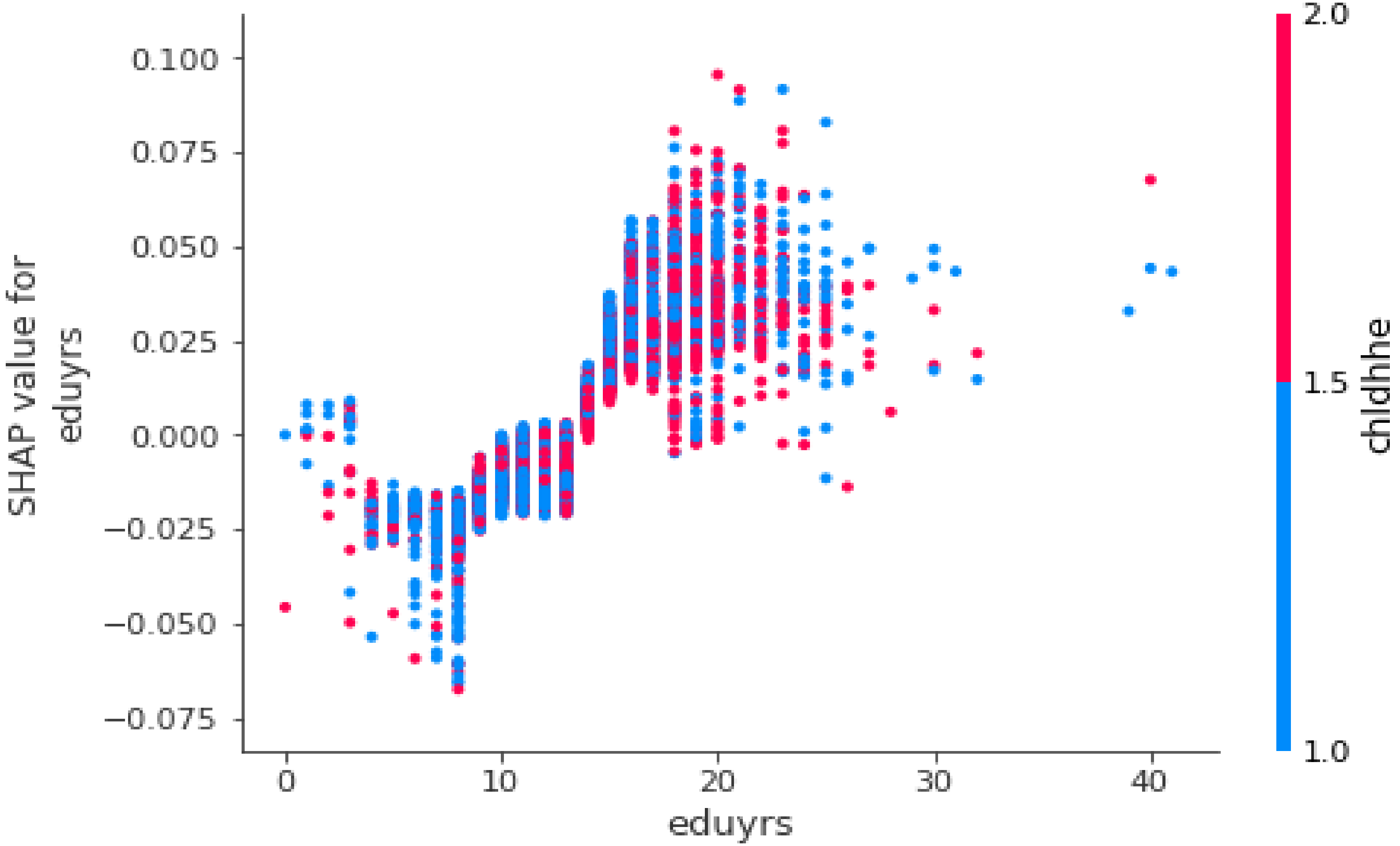


4) EXPLORING THE RELATIONSHIPS AND FORMING HYPOTHESES

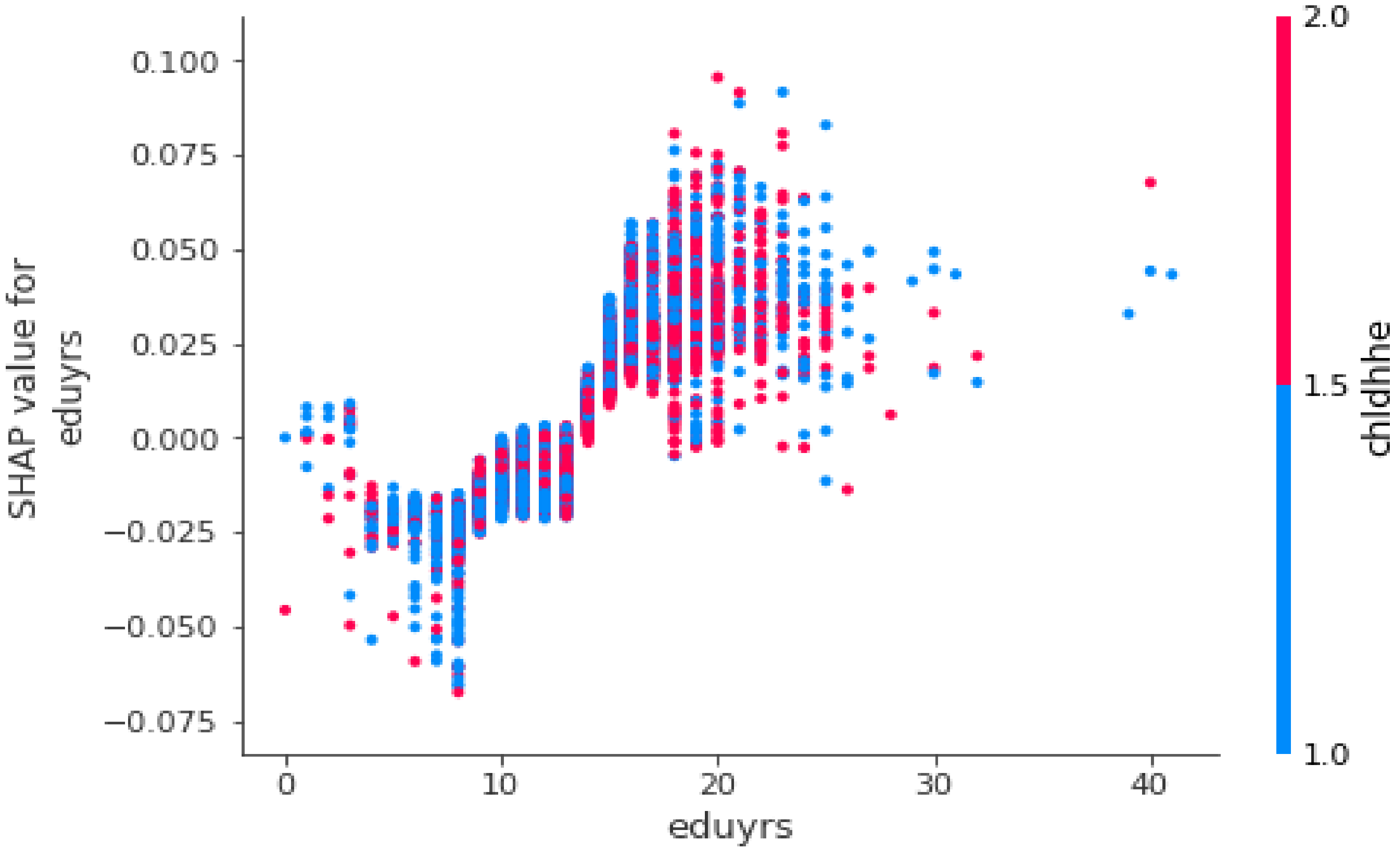


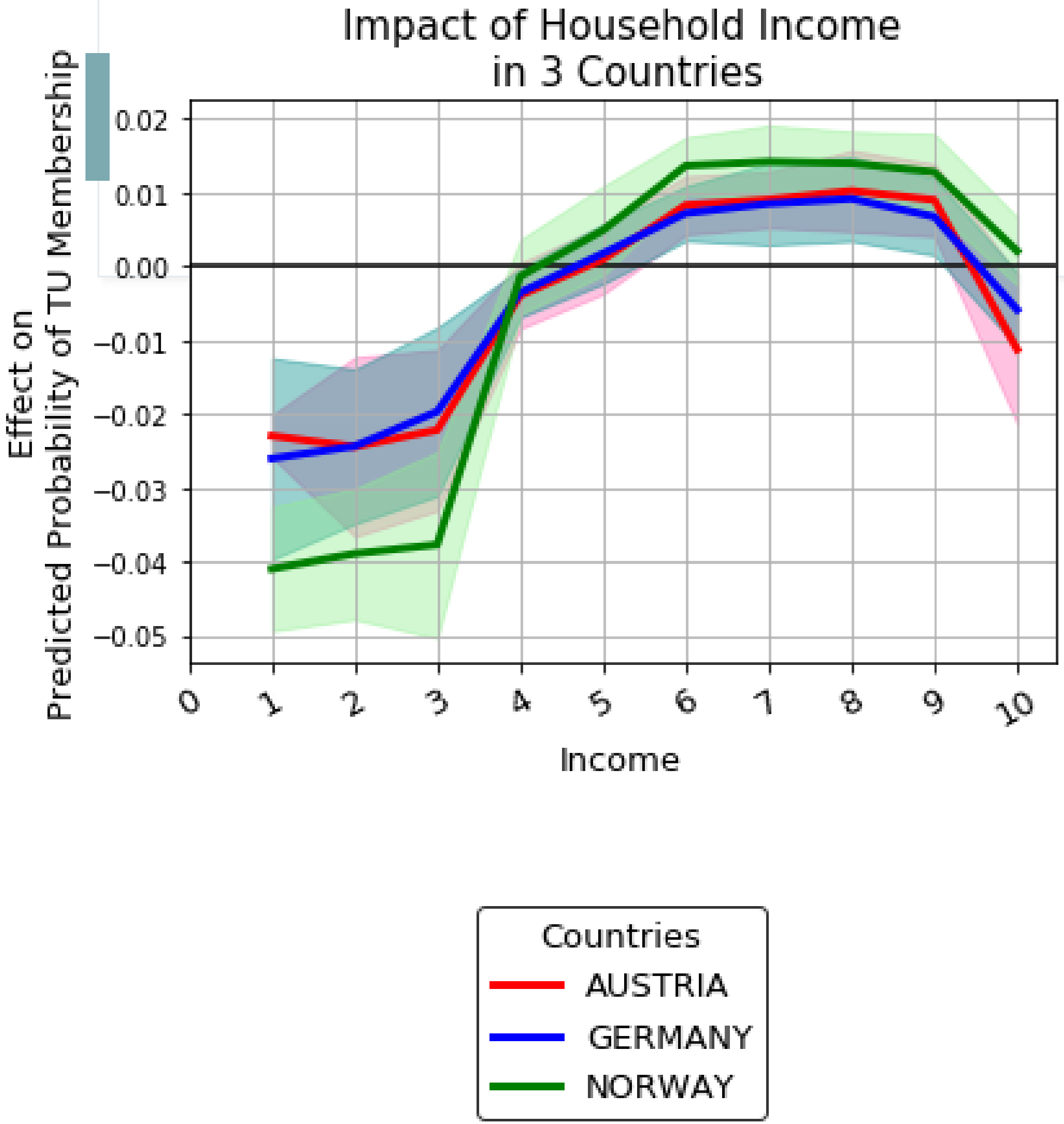


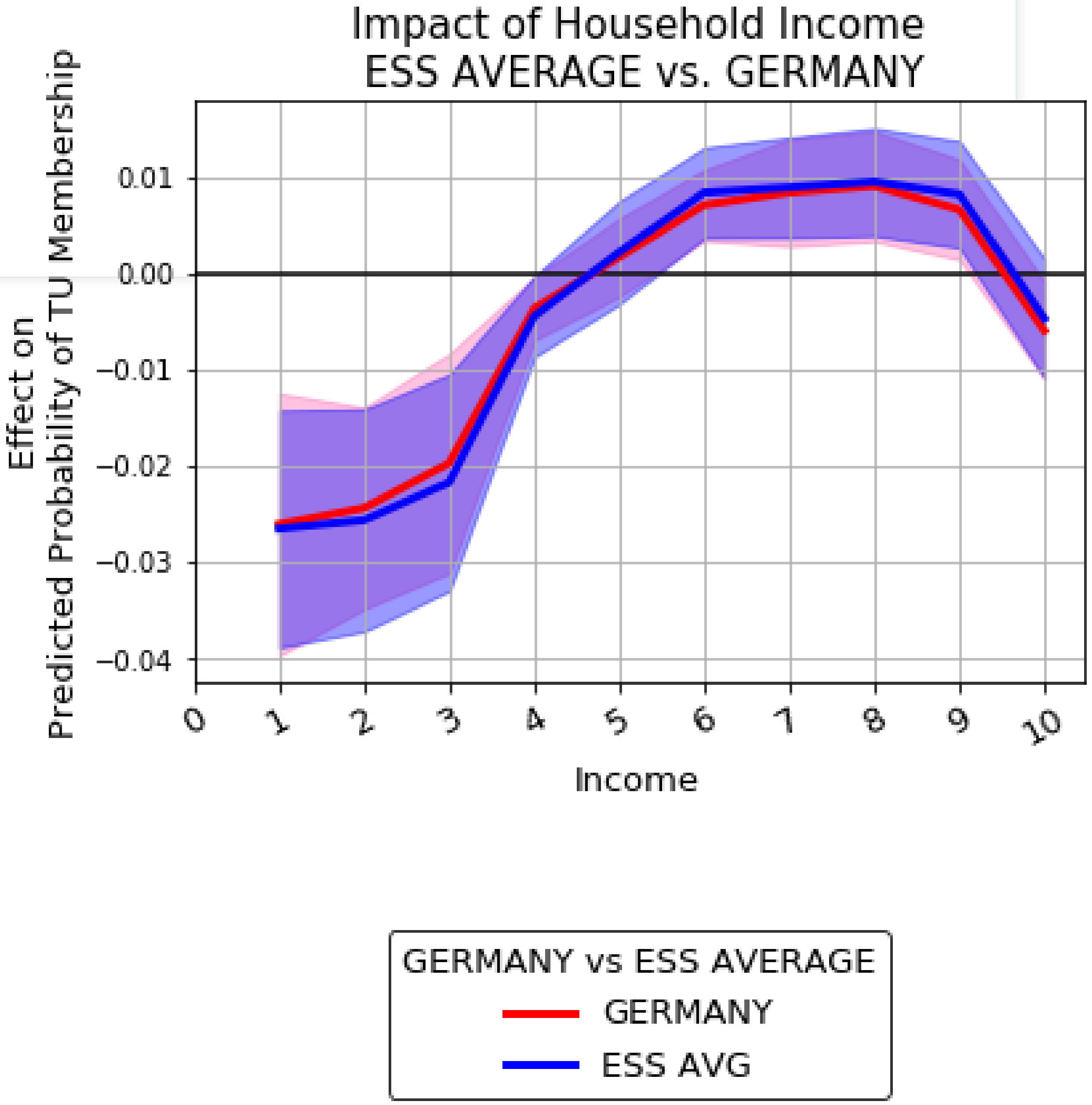
4) EXPLORING THE RELATIONSHIPS AND FORMING HYPOTHESES

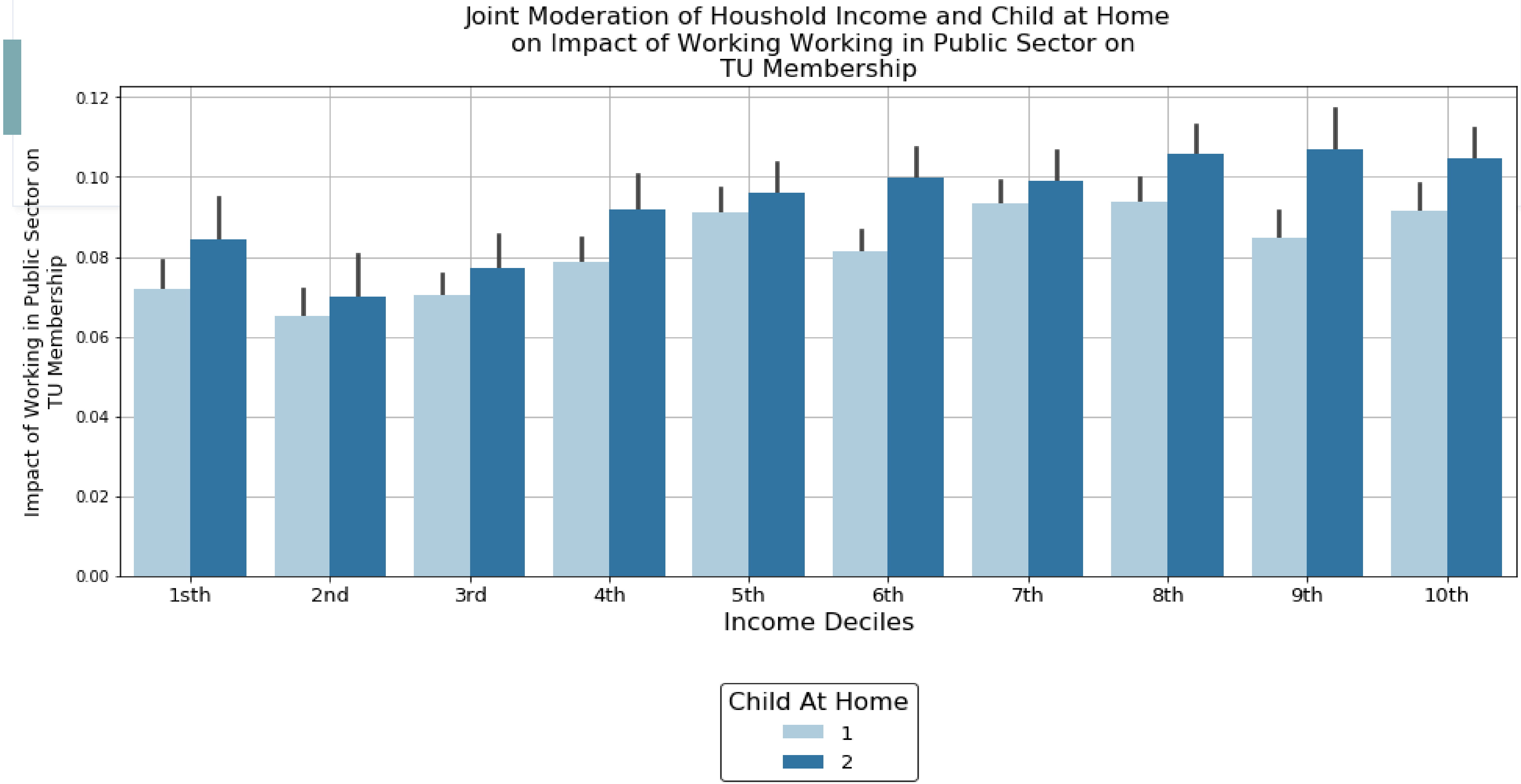


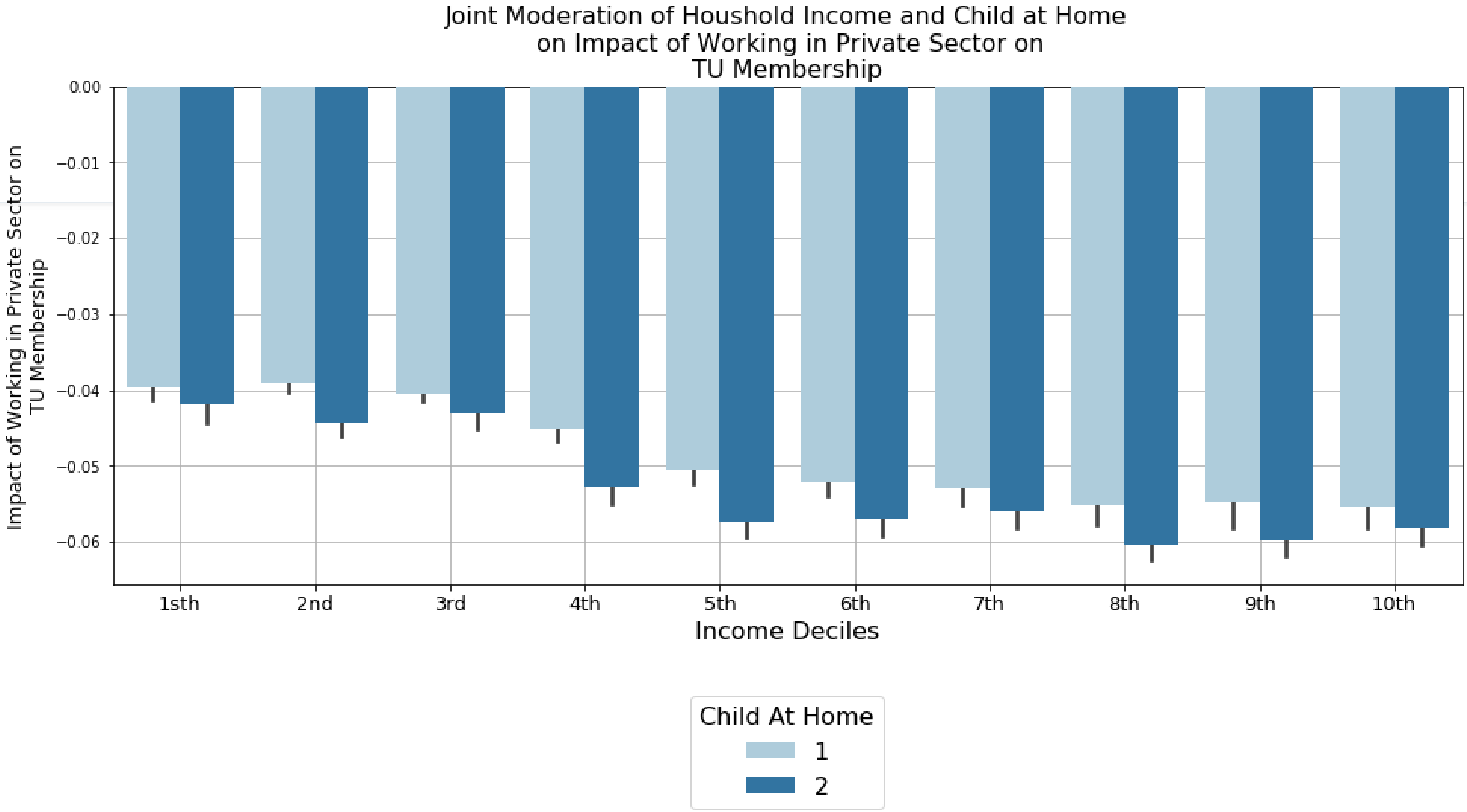
4) EXPLORING THE RELATIONSHIPS AND FORMING HYPOTHESES



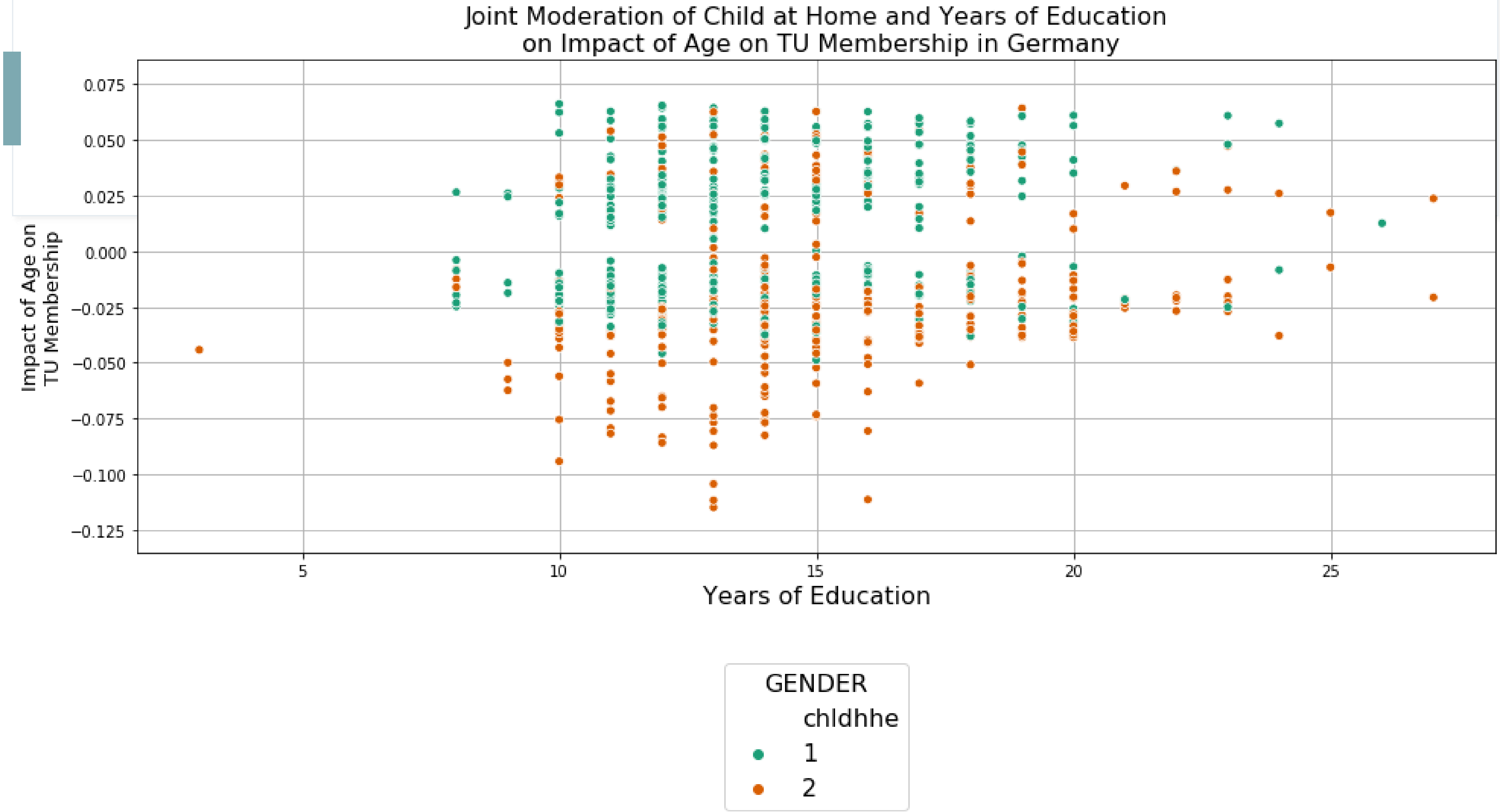

Impact of Household Income
in 3 Countries
Effect on
Predicted Probability of TU Membership
0.02
0.01
0.00
−0.01
−0.02
−0.03
−0.04
−0.05
0
1
2
3
4
5
6
7
8
9
10
Income
Countries
AUSTRIA
GERMANY
NORWAY


Impact of Household Income
ESS AVERAGE vs. GERMANY
Effect on
Predicted Probability of TU Membership
0.01
0.00
−0.01
−0.02
−0.03
−0.04
0
1
2
3
4
5
6
7
8
9
10
Income
GERMANY vs ESS AVERAGE
GERMANY
ESS AVG

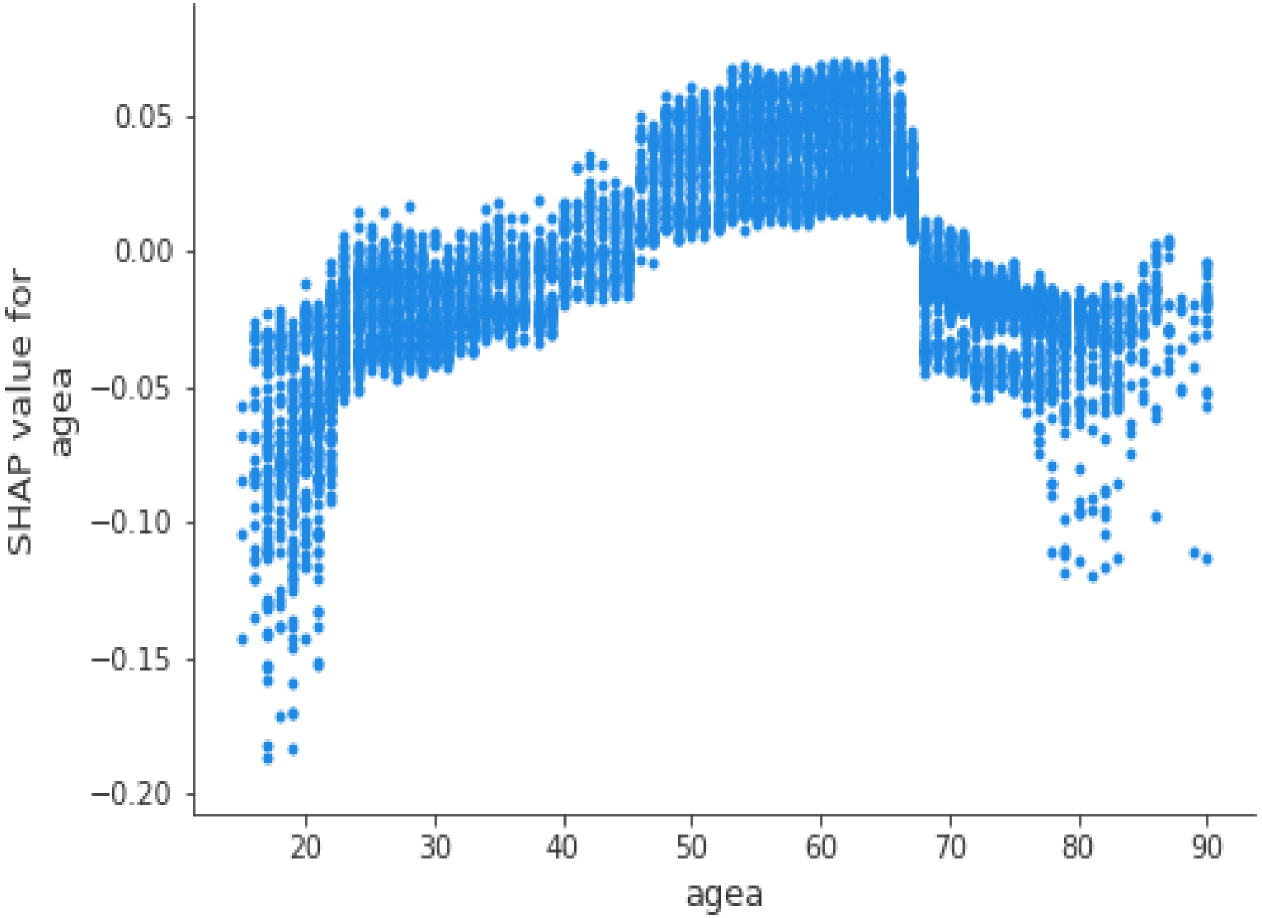

Joint Moderation of Houshold Income and Child at Home
on Impact of Working Working in Public Sector on
TU Membership
Impact of Working in Public Sector on
TU Membership
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1sth
2nd
3rd
4th
5th
6th
7th
8th
9th
10th
Income Deciles
Child At Home
1
2



Joint Moderation of Houshold Income and Child at Home on Impact of Working in Private Sector on TU Membership


Joint Moderation of Child at Home and Years of Education
on Impact of Age on TU Membership in Germany

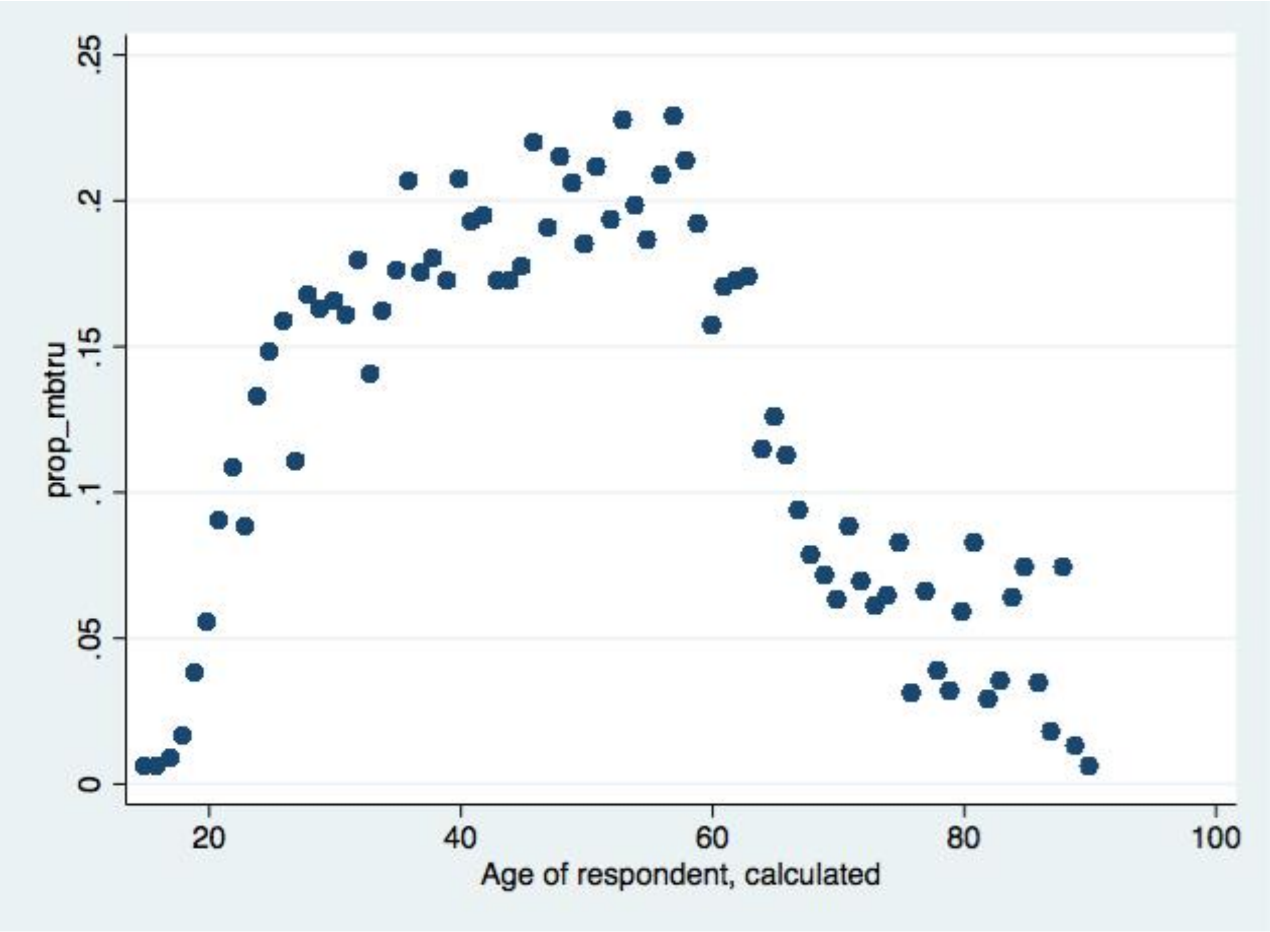
SANITY CHECK



It resembles a 5th order Polynomial

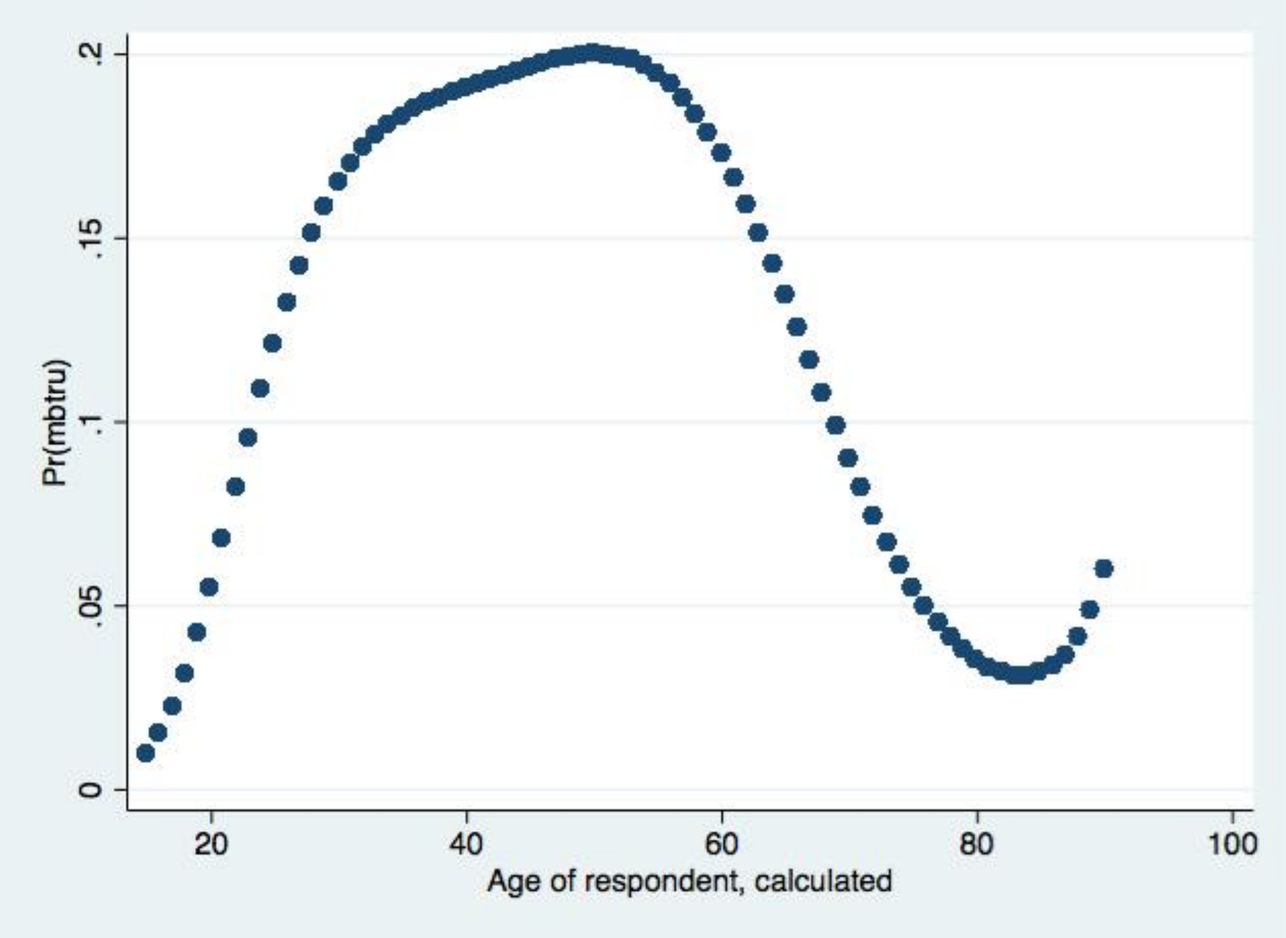
SANITY CHECK

- In order to see the distribution of data I took the mean trade union membership value per each age category. Result is the original distribution.

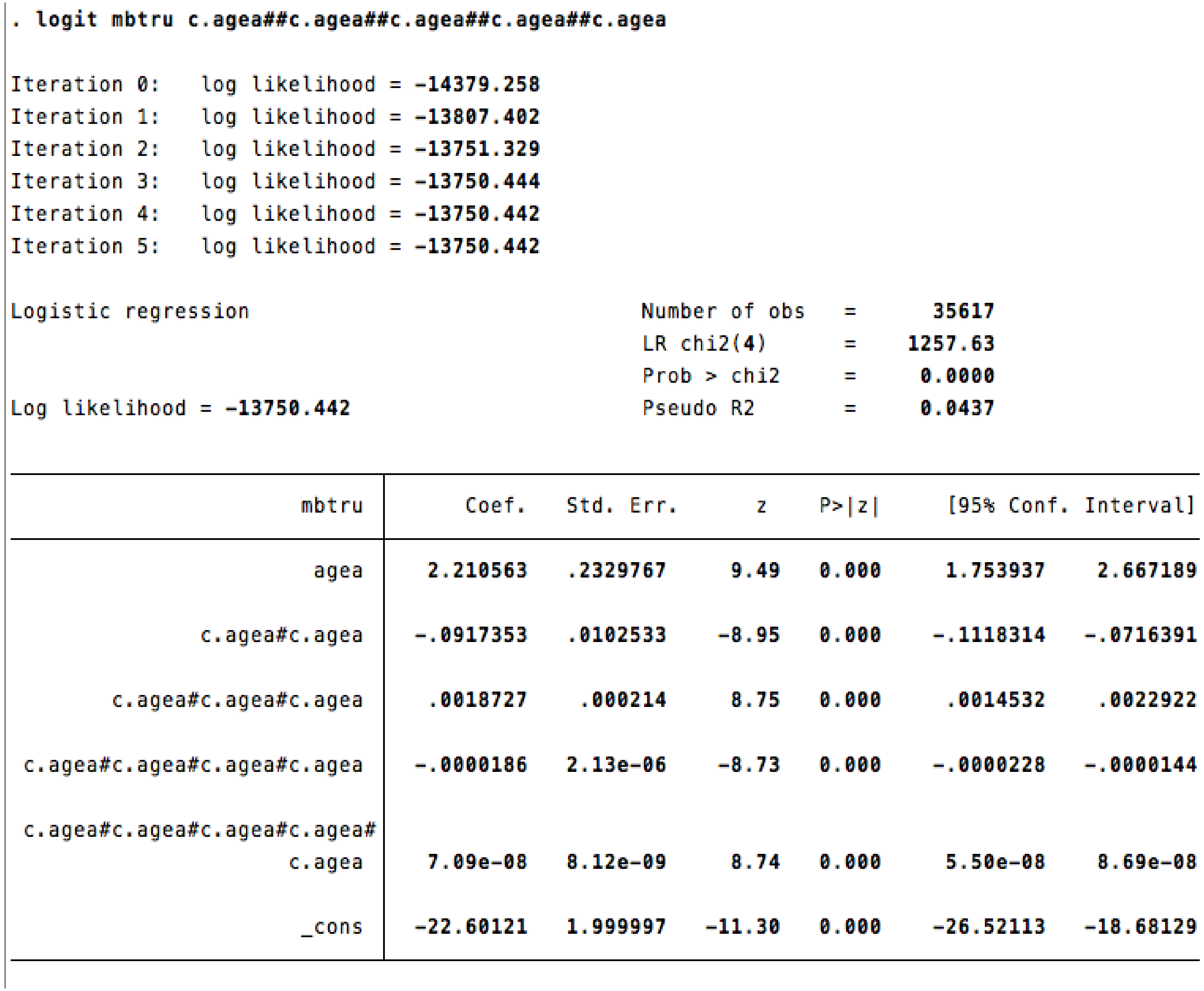


SANITY CHECK

- Then I ran a logistic regression with 5th order age and trade union membership. The plot shows the predictions of that model.



SANITY CHECK: RESULTS



```
. logit mbtru c.agea##c.agea##c.agea##c.agea##c.agea

Iteration 0:   log likelihood = -14379.258
Iteration 1:   log likelihood = -13807.402
Iteration 2:   log likelihood = -13751.329
Iteration 3:   log likelihood = -13750.444
Iteration 4:   log likelihood = -13750.442
Iteration 5:   log likelihood = -13750.442

Logistic regression                             Number of obs     =      35617
                                                LR chi2(4)        =    1257.63
                                                Prob > chi2       =     0.0000
Log likelihood = -13750.442                     Pseudo R2         =     0.0437

------------------------------------------------------------------------------------------
                     mbtru |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------+--------------------------------------------------------------
                      agea |   2.210563   .2329767     9.49   0.000     1.753937    2.667189
                           |
               c.agea#c.agea |  -.0917353   .0102533    -8.95   0.000    -.1118314   -.0716391
                           |
        c.agea#c.agea#c.agea |   .0018727    .000214     8.75   0.000     .0014532    .0022922
                           |
 c.agea#c.agea#c.agea#c.agea |  -.0000186   2.13e-06    -8.73   0.000    -.0000228   -.0000144
                           |
 c.agea#c.agea#c.agea#c.agea#|
                    c.agea |   7.09e-08   8.12e-09     8.74   0.000     5.50e-08    8.69e-08
                           |
                     _cons |  -22.60121   1.999997   -11.30   0.000    -26.52113   -18.68129
------------------------------------------------------------------------------------------
```