Panel Data Analysis of Microeconomic Decisions

Assignment 2 *Mike Weltevrede (1257560)*

1 Binary choice models

1.1

In Table A.1 you can find descriptive statistics for all variables except for ID, YEAR, and WAVE. Some notes that we place with this output:

- There is unbalancedness in the INCOME variable. For all variables, except for JOB_SAT and INCOME, we have 90,056 observations, but for INCOME we have 88,750 observations. For the missing values in INCOME, we tabulate the EMPL variable in Table A.3. We would expect that INCOME is missing when EMPL==0, since perhaps respondents do not fill in INCOME as 0 but rather leave it blank. However, in the output we see that INCOME is only missing if EMPL is equal to 1.
- For JOB_SAT, we have only 48,824 observations. We see in Table A.7 that there are 48,824 observations for EMPL equal to 1. JOB_SAT is missing for all cases where EMPL equal 0. Vice versa, we see in Table A.4 that EMPL can be both 0 and 1 for missing JOB_SAT, but that in 99.14 percent of the cases it is only missing for EMPL equal to 0.
- NUM_CH has a maximum of 11, which seems high. We see that there are indeed quite a few observations for high NUM_CH but that this is minimal in the face of a total of 90,056 observations, especially given that the time aspect (i.e. repetition of the same individual) is not taken into account yet. The results are found in Table A.2.
- SP_INC has a higher maximum, but a lower mean than INCOME. We see that almost 70% has SP_INC equal to 0 (Table A.5) while INCOME is 0 in only 47% of the cases (Table A.6).
- Consider Table A.8. To confirm current societal suspicions, note that mean INCOME for men (SEX equal to 0) is substantially higher than for women (SEX equal to 1) and that mean SP_INC for men is lower than for women, assuming a dominance of heterosexual couples, on average.

1.2

We want to use a static pooled logit model to explain employment. Please find the results in Table A.9. We would like to make several remarks:

- All results are statistically significant at a level $\alpha = 0.001$.
- The variate NUM_CH has a positive coefficient, implying that having more children means there is a higher chance of being employed. This seems partially realistic but also counter-intuitive, seeing as having more children means that you do need more money to sustain them, but you should also need more time to care for them (despite YOUNG_CH being included in the model, this only defines children below the age of 11, and children between ages 12 and 18 still need substantial support from their parents).
- The dummy WHITE has a negative coefficient, referring to a negative effect regardless of the actual value, which seems unlikely in and of itself given possible discriminatory tendencies in society when hiring new people.

1.3

Next, we want to use a static random effects (RE) logit model for the same specification as in the previous question. The results of this can be found in Table A.9. We see that all effects become more pronounced, i.e. the estimated coefficient is larger in the direction of its sign. This is because the pooled logit model did not take individual effects into account, whereas the RE logit model does. The former led to an underestimation due to individual effects having to be included in the error term, seeing as they most certainly exist in a panel data set. In the next exercise, we will determine whether these individual effects are substantial enough.

1.4

In this exercise, we want to know whether the pooled logit or the RE logit model is more appropriate. Firstly, some intuition. Consider Table A.9. Note that $\rho = 71.5\%$, so there is a large portion of unobserved heterogeneity in the individual effects. As such, it is likely that the RE logit model will be proven to be the better model over the pooled logit model, which does not take the panel data structure into account.

For the test, we can do a likelihood ratio test, seeing as RE logit is a more restrictive version of the pooled logit model. This gives us the hypotheses:

$$H_0 : \text{pooled logit model} \quad \text{vs.} \quad H_1 : \text{random effects model}.$$

Notice that the log-likelihoods are $LL_{pooled} = -49944.725$ and $LL_{RE} = -36144.325$, as retrieved from Table A.9, yielding the test statistic:

$$LR = 2\left(LL_{RE} - LL_{pooled}\right) = 2\left(-36144.325 + 49944.725\right) = 27600.80$$

The $\chi^2_{1;0.95}$ critical value is 3.84, so this realisation of the test statistic exceeds it. As such, we reject H_0 in favour of H_1, i.e. the RE model is to be preferred.

1.5

We want to investigate whether there is a non-monotonic effect of age on employment status using a formal test. For this, we create a quadratic age term and add it to the RE model selected in the previous exercise. The output can be found in Table A.9. We can immediately see that the estimated coefficients apart from those for AGE and AGE2 do not change much. This is the case because those coefficients were already estimated controlling for AGE, so the added effect of AGE2 has little effect. Regarding the non-monotonic effect of AGE, we want to do a test with the following hypotheses:

$$H_0 : \beta_{AGE2} = 0 \quad \text{vs.} \quad H_1 : \beta_{AGE2} \neq 0.$$

We see that AGE2 has a p-value of 0.035, so we find statistical evidence at a significance level of 0.05 that there is a non-monotonic effect of age on employment status, ceteris paribus. As such, we reject H_0 in favour of H_1 at a significance level of $\alpha = 0.05$. We can also see this since 0 is not included in the 95% confidence interval for $\hat{\beta}_{AGE2}$. Note also that the signs of the estimates differ, as $\hat{\beta}_{AGE} = -0.035$ and $\hat{\beta}_{AGE2} = 0.009$.

1.6

We want to compute the marginal effect of a one year increase in age on the probability to be employed for an observation with probability 0.5 to be employed and age 40 years old. Then, this leads to the following derivation:

$$\begin{aligned}\frac{\partial}{\partial AGE}\mathbb{P}\left[EMPL = 1 \mid x_{it}, \alpha_i\right] &= (\beta_{AGE} + 2\beta_{AGE2} \cdot AGE) \cdot \frac{\partial}{\partial x_{it}'\beta}\Lambda(x_{it}'\beta + \alpha_i) \\ &= (\beta_{AGE} + 2\beta_{AGE2} \cdot AGE) \cdot \Lambda(x_{it}'\beta + \alpha_i) \cdot (1 - \Lambda(x_{it}'\beta + \alpha_i)) \quad (1)\end{aligned}$$

Evaluating (1) at the point estimates that follow from the model and using that $x_{it}'\beta + \alpha_i = 0.5$, we get:

$$\begin{aligned}\left(\hat{\beta}_{AGE} + 2\hat{\beta}_{AGE2}\cdot AGE\right)\cdot\Lambda(x_{it}'\beta+\alpha_i)\cdot(1-\Lambda(x_{it}'\beta+\alpha_i)) &= \left(\hat{\beta}_{AGE} + 2\hat{\beta}_{AGE2}\cdot AGE\right)\cdot 0.5\cdot(1-0.5)\\ &= (-0.1313128 + 2\cdot 0.0008531\cdot 40)\cdot 0.25\\ &\approx -0.0158\end{aligned}$$

We see that there is a negative effect of age on employment status, namely that the probability of being employed at age 41, ceteris paribus, drops by 0.0158, meaning that the person in question has a probability of being employed of $0.5 - 0.0158 = 0.4842$ at age 41.

We are now interested in knowing for which age the marginal effect is 0. Using Equation (1) and the point estimates, we get:

$$\begin{aligned}&\left(\hat{\beta}_{AGE} + 2\hat{\beta}_{AGE2}\cdot AGE\right)\cdot\Lambda(x_{it}'\beta+\alpha_i)\cdot(1-\Lambda(x_{it}'\beta+\alpha_i)) &&= 0\\ \Longleftrightarrow\ &\hat{\beta}_{AGE} + 2\hat{\beta}_{AGE2}\cdot AGE &&= 0\\ \Longleftrightarrow\ &AGE &&= -\frac{\hat{\beta}_{AGE}}{2\hat{\beta}_{AGE2}}\end{aligned}$$

Again, using the point estimates, this evaluates to $\frac{-0.1313128}{2\cdot 0.0008531} \approx 76.96$. So, roughly at age 77, the marginal effect is 0.

1.7

For this exercise, we estimated a static fixed effects (FE) model using the specification without AGE2. The results are found in Table A.9. Firstly, some comments on the results:

- WHITE is excluded due to transformations removing time-invariant regressors.
- The estimates for β are generally less pronounced (except for EDU_15) and with higher standard errors, as we can expect. However, note that the standard error for $\hat{\beta}_{MARRIED}$ is lower for the FE model than for the RE model. A Hausman test on this variable would be inconclusive.

Now, we want to do a test to choose between the FE and RE models, i.e. we have the hypotheses:

$$H_0 : \text{random effects model} \quad \text{vs.} \quad H_1 : \text{fixed effects model.}$$

We do a Hausman test but we cannot use all coefficients so we only use one coefficient of interest which we choose a priori. Since before we were interested in the effect of age, we will look at AGE as our variable of interest. We then have the following test statistic:

$$T_{AGE} = \frac{\hat{\beta}_{FE,AGE} - \hat{\beta}_{RE,AGE}}{\sqrt{s.e.^2_{FE,AGE} - s.e.^2_{RE,AGE}}} = \frac{-0.0626 - (-0.0637)}{\sqrt{0.0027^2 - 0.0025^2}} \approx 1.0786.$$

Then $|T| = 1.0786 < t_{0.975} = 1.96$ so we do not find sufficient evidence at $\alpha = 0.05$ (note that we do a two-tailed t-test) to reject H_0. As such, the RE model is to be preferred over the FE model.

1.8

We now want to estimate a quasi fixed effects (QFE) model that allows the individual effects to be correlated with education. Again, the results are found in Table A.9. We constructed this model by adding means over time for the education dummies, which we call MEAN_EDU_13_15 and MEAN_EDU_15. We make the following remarks:

- Both time-means are significant at a level of $\alpha = 0.1$, with MEAN_EDU_13_15 being significant at $\alpha = 0.05$. This variate has a positive coefficient, which indicates that there is a positive effect on employment status.
- Compared to the RE model, the results are less pronounced (except for MARRIED and EDU_15).
- Note that the standard errors, indicating precision, are generally the same when comparing the QFE model to the RE model. This is due to a similar reasoning we used when adding AGE2 to the model, namely that the estimates for the non-education regressors were already estimated taking EDU_13_15 and EDU_15 into account, so their precision does not change much. For the education dummies, however, we see a higher standard error, so a lower precision, due to correlation with the newly added time-means.

1.9

We want to test the QFE model versus the FE model. Note that QFE is a special case of RE (i.e. it is consistent and efficient under the null hypothesis that there are significant individual effects) and that we can therefore do a Hausman test. We have:

$$H_0 : \text{quasi fixed effects model} \quad \text{vs.} \quad H_1 : \text{fixed effects model.}$$

This time, we look at the education dummies since these are the reason our QFE model exists. Consider EDU_13_15 as our variable of interest. We then have the following test statistic:

$$T_{AGE} = \frac{\hat{\beta}_{FE,EDU\_13\_15} - \hat{\beta}_{QFE,EDU\_13\_15}}{\sqrt{s.e.^2_{FE,EDU\_13\_15} - s.e.^2_{QFE,EDU\_13\_15}}} = \frac{-0.0626 - (-0.0639)}{\sqrt{0.0027^2 - 0.0026^2}} \approx 1.7857.$$

Then $|T| = 1.7857 < t_{0.975} = 1.96$ so we do not find sufficient evidence at $\alpha = 0.05$ to reject H_0. As such, the QFE model is to be preferred over the FE model.

1.10

In this last subquestion, we are interested in a dynamic RE model to see whether there is state dependence using the Wooldridge approach. Recall that the Wooldridge approach models $\alpha_i = \lambda y_{i1} + \tilde{\alpha}_i$ and, as such, it is simply a random effects model where the initial value of y, in our case this is EMPL, is added to the model. The results are found in Table A.9. We can see that the estimate for λ is significantly different from zero at significance level $\alpha = 0.01$. As such, there is indeed state dependence on the initial employment status, meaning there is correlation between the personal characteristics α_i and EMPL. Moreover, note that there is significant unobserved heterogeneity, supported by the estimate for σ_α being statistically significant at $\alpha = 0.01$.

2 Tobit models

2.1

We will estimate a static RE tobit model to explain INCOME from SEX, AGE, WHITE, MARRIED, and the education dummies. The results can be found in Table A.10. Some interesting results are:

- Again, note that the estimate for the coefficient for WHITE is negative, meaning that a white person would be expected to earn less than a non-white person. Using the same analysis as before, this seems not intuitive in current day's society.
- Education gives, as expected, a significantly higher expected income the more schooling is followed.
- A married person earns a higher income than an unmarried person. This may be because their spouse could take more care of the household that the person themselves otherwise may have had to do themselves, allowing this person to work more hours.

2.2

We can see that there is an estimate for ρ in the model output. Its value is 0.6513, i.e. 65.13% of the unobserved heterogeneity is caused by the individual effects as compared to 34.87% from the error terms. As such, there is for sure a strong individual effect through the alphas.

2.3

We are interested in knowing whether the effect of marital status on income differs for males and females. To this end, we add the interaction term between MARRIED and SEX to the model from question 2.1. The results can also be found in Table A.10. We see that the estimate for $\beta_{MARRIED\_SEX}$ is significantly different from zero at significance level $\alpha = 0.01$. As such, we find statistically significant evidence that there is a different effect of marital status on income for males and females. Since the coefficient is negative, there is a less strong effect, as the effect of marital status for a married man is 36692.8, while an additional 17413.32 gets subtracted for a married woman, leading to an effect of $36692.8 - 17413.32 = 19279.48$. Note that this is not the marginal effect. For that, we still need to do some extra calculations.

2.4

We want to estimate the dynamic version of the model estimated in question 2.1 using the Wooldridge approach (as also used in 1.10). The results can be found in Table A.10. Interestingly, the effects of SEX and the education dummies become less pronounced when taking the initial condition into account. All other effects stay roughly the same. We can see that the importance of unobserved heterogeneity, explained by ρ, also decreases. It was 65.13% in the static model, whereas it is 52.40% in the dynamic model. Nonetheless, this is a substantial proportion.

2.5

To test whether the individual effects are correlated with the initial value of INCOME, we want to know whether the estimate for λ is statistically different from zero. In Table A.10, we see that it is indeed statistically significant at $\alpha = 0.01$. As such, we determine that there is correlation between the individual effects and the initial value of INCOME.

3 Ordered response models

3.1

In this question, we want to estimate a static random effects ordered probit model to explain job satisfaction (JOB_SAT) from age, income and education. In this case we are only interested in the sub-sample of employed individuals. The results from this model are found in Table A.11. We see that there is no statistical evidence to suggest that the variable EDU_13_15 has an effect on JOB_SAT. However, all other variables do have a statistically significant (negative) effect. This is logical seeing as JOB_SAT is measured from on a scale from 1 (like it very much) to 4 (dislike it very much). As such, a higher INCOME intuitively leads to a higher satisfaction with one's job, for example.

3.2

Note that the output for the model from question 3.1 does not contain an estimate for the constant term. This is because we need to take care of the identification problem when estimating our parameters since the cut-off points also need to be estimated. This can be done by fixing the intercept or the first cut-off point to be 0 for normalisation.

3.3

At the bottom of the Stata output for the ordered probit model, we see the following likelihood ratio test being carried out:

```
LR test vs. oprobit regression:  chibar2(01) =  6936.68 Prob>=chibar2 = 0.0000
```

This is a test for the following hypotheses:

$$H_0 : \text{pooled ordered probit model} \quad \text{vs.} \quad H_1 : \text{non-pooled ordered probit model.}$$

We can see this as the test is "vs. oprobit regression", where `oprobit` is the pooled version of `xtoprobit`, i.e. that there is no statistical evidence for the influence of individual effects, i.e. there is no evidence that ρ is different from zero. In this case, we see that we reject H_0 and that there are indeed individual effects and that a panel data structure should be assumed.

3.4

We now want to compute some marginal effects. Consider the following:

- What is the marginal effect of EDU_15 on the probability that JOB_SAT=1 for an observation with probability of JOB_SAT=1 equal to 0.5?

$$\begin{aligned}
\frac{\partial}{\partial x_{i,EDU_15}} \mathbb{P}[JOB_SAT = 1 \mid x_i] &= \frac{\partial}{\partial x_{i,EDU_15}} \mathbb{P}[JOB_SAT^* < m_0 \mid x_i] \\
&= \frac{\partial}{\partial x_{i,EDU_15}} \Phi(m_0 - (x_i'\beta + \alpha_i)) \\
&= -\beta_{EDU_15} \cdot \phi(m_0 - (x_i'\beta + \alpha_i)) \\
&= -\beta_{EDU_15} \cdot \phi\left(\Phi^{-1}(0.5)\right),
\end{aligned}$$

  where $\phi\left(\Phi^{-1}(0.5)\right) = \phi(0) = (2\pi)^{-\frac{1}{2}} \approx 0.3989$. Evaluating this at the point estimate $\hat{\beta}_{EDU\_15}$ gives: $-(-0.1866) \cdot 0.3989 \approx 0.0744$.

- What is the marginal effect of EDU_15 on the probability that JOB_SAT=4 for an observation with probability of JOB_SAT=4 equal to 0.5?

$$\begin{aligned}
\frac{\partial}{\partial x_{i,EDU_15}} \mathbb{P}[JOB_SAT = 4 \mid x_i] &= \frac{\partial}{\partial x_{i,EDU_15}} \mathbb{P}[JOB_SAT^* > m_3 \mid x_i] \\
&= \frac{\partial}{\partial x_{i,EDU_15}} \Phi(x_i'\beta + \alpha_i - m_3) \\
&= \beta_{EDU_15} \cdot \phi(x_i'\beta + \alpha_i - m_3) \\
&= \beta_{EDU_15} \cdot \phi(\Phi^{-1}(0.5))
\end{aligned}$$

  Evaluating this at the point estimate $\hat{\beta}_{EDU\_15}$ gives: $-0.1866 \cdot 0.3989 \approx -0.0744$.

We see that these outcomes are the same apart from the sign, simply due to the probabilities of the specific JOB_SAT values being the same. As such, the marginal effects of EDU_15 can be interpreted as such: someone with highest degree being graduate level is, compared to someone with highest degree being high school level, 7.44% more c.q. less likely to have JOB_SAT equal to 1 c.q. 4.

A Tables

Table A.1: Descriptive statistics, excluding ID, YEAR, and WAVE

Variable	Obs	Mean	Std. Dev.	Min	Max
AGE	90056	39.64884	5.111332	29	51
WHITE	90056	.5751977	.4943157	0	1
MARRIED	90056	.3976082	.4894063	0	1
EMPL	90056	.5460713	.4978757	0	1
YOUNG_CH	90056	.4590588	.4983238	0	1
EDU_12	90056	.5634272	.4959634	0	1
EDU_13_15	90056	.2250267	.4176022	0	1
EDU_15	90056	.2115461	.4084069	0	1
SEX	90056	.510971	.4998824	0	1
INCOME	88750	19769.22	33024.04	0	307823
SP_INC	90056	12459.57	29250.6	0	309409
NUM_CH	90056	1.396564	1.36329	0	11
JOB_SAT	48824	1.606218	.6821562	1	4

Table A.2: Tabulating NUM_CH

NUM_CH	Freq.	Percent	Cum.
0	32,897	36.53	36.53
1	15,467	17.17	53.70
2	23,535	26.13	79.84
3	12,163	13.51	93.34
4	4,196	4.66	98.00
5	1,199	1.33	99.33
6	371	0.41	99.75
7	142	0.16	99.90
8	54	0.06	99.96
9	15	0.02	99.98
10	10	0.01	99.99
11	7	0.01	100.00
Total	90,056	100.00	

Table A.3: Tabulating EMPL for missing INCOME

EMPL	Freq.	Percent	Cum.
1	1,306	100.00	100.00
Total	1,306	100.00	

Table A.4: Tabulating EMPL for missing JOB_SAT

EMPL	Freq.	Percent	Cum.
0	40,879	99.14	99.14
1	353	0.86	100.00
Total	41,232	100.00	

Table A.5: Tabulating SP_INC. Value 1 corresponds to any nonzero SP_INC.

SP_INC_TAB	Freq.	Percent	Cum.
0	62,416	69.31	69.31
1	27,640	30.69	100.00
Total	90,056	100.00	

Table A.6: Tabulating INCOME. Value 1 corresponds to any nonzero INCOME.

INCOME_TAB	Freq.	Percent	Cum.
0	42,123	47.46	47.46
1	46,627	52.54	100.00
Total	88,750	100.00	

Table A.7: Tabulating JOB_SAT for EMPL equal to 1

JOB_SAT	Freq.	Percent	Cum.
Like it very much	23,803	48.75	48.75
Like it fairly well	21,399	43.83	92.58
Dislike it somewhat	2,667	5.46	98.04
Dislike it very much	955	1.96	100.00
Total	48,824	100.00	

Table A.8: Summarising INCOME and SP_INC split on SEX.

Variable	SEX	Obs	Mean	Std. Dev.	Min	Max
INCOME	0 (Male)	43,417	25251.48	39845.78	0	307823
	1 (Female)	45,333	14518.67	23625.57	0	307823
SP_INC	0 (Male)	44,040	7742.282	19217.31	0	309409
	1 (Female)	46,016	16974.29	35767.94	0	309409

Table A.9: Model outputs for Question 1 - Binary Choice Models

EMPL	Static Pooled Logit	Static RE Logit	Static RE Logit	Static FE Logit	Static QFE Logit	Dynamic RE Logit
Intercept	1.2656***	2.2523***	3.5650***		2.2308***	-0.0458
	(0.0621)	(0.1110)	(0.6334)		(0.1144)	(0.1351)
AGE	-0.0359***	-0.0637***	-0.1313***	-0.0626***	-0.0639***	-0.0649***
	(0.0015)	(0.0025)	(0.0322)	(0.0027)	(0.0026)	(0.0031)
MARRIED	2.2026***	2.7174***	2.7176***	1.7476***	2.7181***	2.5181***
	(0.0181)	(0.0393)	(0.0393)	(0.0384)	(0.0393)	(0.0415)
YOUNG_CH	-0.8762***	-1.1566***	-1.1583***	-0.9564***	-1.1562***	-1.1629***
	(0.0198)	(0.0356)	(0.0356)	(0.0363)	(0.0356)	(0.0382)
NUM_CH	0.0979***	0.2253***	0.2281***	0.2167***	0.2242***	0.2615***
	(0.0074)	(0.0200)	(0.0201)	(0.0289)	(0.0201)	(0.0197)
EDU_13_15	0.3606***	0.6816***	0.6812***	0.5183***	0.5102***	0.3689***
	(0.0193)	(0.0625)	(0.0626)	(0.1060)	(0.1117)	(0.0599)
EDU_15	0.4404***	1.2225***	1.2231***	1.4845***	1.4969***	0.6110***
	(0.0204)	(0.0739)	(0.0740)	(0.1752)	(0.1831)	(0.0689)
WHITE	-0.6065***	-1.1127***	-1.1126***	0.0000	0.0000	-0.7652***
	(0.0161)	(0.0622)	(0.0622)	(.)	(.)	(0.0560)
AGE2			0.0009**			
			(0.0004)			
MEAN_EDU_13_15					0.3309**	
					(0.1361)	
MEAN_EDU_15					-0.3471*	
					(0.1992)	
EMPL_INIT						3.9348***
						(0.0639)
sigma_u		2.8755***	2.8757***		2.8743***	2.3281***
		(0.0362)	(0.0362)		(0.0362)	(0.0317)
rho		0.7154***	0.7154***		0.7152***	0.6223***
		(0.0051)	(0.0051)		(0.0051)	(0.0064)
Log-likelihoods	-49944.725	-36144.325	-36142.108	-15697.873	-36136.269	-29104.543

***, **, and * indicate significance at significance levels 0.01, 0.05, and 0.1, respectively. Standard errors in parentheses.

Table A.10: Model outputs for Question 2 - Tobit Models

INCOME	Static RE Tobit	Static RE Tobit	Dynamic RE Tobit
Intercept	-48620.9***	-51940.93***	-71471.38***
	(1284.4275)	(1283.4220)	(1496.1442)
SEX	-16310.26***	-8456.277***	-3083.767***
	(814.7245)	(873.1604)	(697.4258)
AGE	1041.6285***	1032.6780***	1047.5795***
	(26.0181)	(25.9181)	(32.1262)
WHITE	-11804.21***	-11536.08***	-12504.71***
	(828.2641)	(814.2254)	(689.3718)
MARRIED	28100.31***	36692.8***	28300.87***
	(400.1924)	(556.8055)	(426.8352)
EDU_13_15	10053.59***	9973.3746***	3889.5526***
	(749.8609)	(741.7986)	(707.1805)
EDU_15	27432.88***	27326.62***	10480.56***
	(908.2875)	(895.7245)	(832.4009)
MARRIED_SEX		-17413.32***	
		(776.3575)	
INCOME_INIT			1.5359***
			(0.0196)
sigma_u	39673.22***	38952.7***	30731.77***
	(368.1049)	(363.6634)	(304.7231)
sigma_e	29026.55***	28936.58***	29287.85***
	(103.8570)	(103.5161)	(113.2414)
rho	0.6513	0.6444	0.5240
	(0.0043721)	(0.0044)	(0.0051)
Log-likelihoods	-571545.03	-571294.75	-485393.24

***, **, and * indicate significance at significance levels 0.01, 0.05, and 0.1, respectively. Standard errors in parentheses.

Table A.11: Model outputs for Question 3 - Ordered Choice Models

JOB_SAT	Static Ordered Probit
AGE	-0.0066***
	(0.0013)
INCOME	-1.96×10^{-6} ***
	(0.0000)
EDU_13_15	-0.0371
	(0.0242)
EDU_15	-0.1866***
	(0.0272)
cut1	-0.4090***
	(0.0511)
cut2	1.4727***
	(0.0516)
cut3	2.2250***
	(0.0532)
sigma2_u	0.6542***
	(0.0178)
Log-likelihoods	-41276.021

***, **, and * indicate significance at significance levels 0.01, 0.05, and 0.1, respectively. Standard errors in parentheses.

B Code

```
clear
est clear
set more off
set mem 10m

cd "C:\Users\mikew\Desktop\University\TilburgUniversity\Master\panel_data_analysis\
    nonlinear_models\assignment"
global datapath "data"

use "$datapath\data.dta", clear // Load data
keep if WAVE <= 8 // My SNR is 1257560, so I use waves 1 to 8

// Binary Choice Models
// 1.
sum

// Checking unbalancedness in the INCOME and JOB_SAT variables
tab EMPL if missing(INCOME)

tab JOB_SAT if EMPL==1
tab JOB_SAT if EMPL==0 // no observations

tab EMPL if missing(JOB_SAT)

// High number of children; check distribution of NUM_CH:
tab NUM_CH

// Distribution of 0 INCOME/SP_INC:
gen SP_INC_TAB = 0
replace SP_INC_TAB = 1 if SP_INC > 0
tab SP_INC_TAB

gen INCOME_TAB = 0
replace INCOME_TAB = 1 if INCOME > 0
tab INCOME_TAB if !missing(INCOME)

// Income of men versus women:
sum INCOME if SEX==1
sum INCOME if SEX==0
sum SP_INC if SEX==1
sum SP_INC if SEX==0

// 2.
global predictors_bc AGE MARRIED YOUNG_CH NUM_CH EDU_13_15 EDU_15 WHITE

logit EMPL $predictors_bc
est store pooled_logit

// 3.
xtset ID WAVE
xtlogit EMPL $predictors_bc , re
est store re_logit

// 5.
gen AGE2 = AGE^2
xtlogit EMPL $predictors_bc AGE2, re
est store re_logit_age_squared

// 7.
xtlogit EMPL $predictors_bc , fe
est store fe_logit

// 8.
```

```
egen MEAN_EDU_13_15 = mean(EDU_13_15), by(ID)
egen MEAN_EDU_15 = mean(EDU_15), by(ID)

xtlogit EMPL $predictors_bc MEAN_EDU_13_15 MEAN_EDU_15 WHITE, re
est store qfe_logit

// 10.
gen EMPL_INIT = EMPL
by ID, sort: replace EMPL_INIT = EMPL[1]

xtlogit EMPL $predictors_bc EMPL_INIT if WAVE > 1, re
est store dynamic_re_logit

estout *,cells(b(star fmt(%9.4f)) se(par fmt(%9.4f))) style(smcl) ///
starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) modelwidth(12)

// Tobit Models
est clear

// 1.
global predictors_tobit SEX AGE WHITE MARRIED EDU_13_15 EDU_15
xttobit INCOME $predictors_tobit, ll(0)
est store tobit_out

// 3.
gen MARRIED_SEX = MARRIED*SEX
xttobit INCOME $predictors_tobit MARRIED_SEX, ll(0)
est store tobit_sex_married

// 4.
gen INCOME_INIT = INCOME
by ID, sort: replace INCOME_INIT = INCOME[1]

xttobit INCOME $predictors_tobit INCOME_INIT if WAVE > 1, ll(0)
est store dynamic_tobit

estout *,cells(b(star fmt(%9.4f)) se(par fmt(%9.4f))) style(smcl) ///
starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) modelwidth(12)

// Ordered Response Models
est clear

// 1.
xtoprobit JOB_SAT AGE INCOME EDU_13_15 EDU_15 if EMPL==1
est store ordered_probit

estout *,cells(b(star fmt(%9.4f)) se(par fmt(%9.4f))) style(smcl) ///
starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) modelwidth(12)
```