Assignment 2:
Panel Data Analysis of Microeconomic Decisions

Daniel Redel

Fall 2022

1 Binary Choice Models

Table 1: Binary Choice Models Regression Results

	(1)	(2)	(3)	(4)	(5)
Employed	Pooled Logit	RE Logit	FE Logit	QFE Logit	Dynamic RE
Age	-0.071***	-0.131***	-0.122***	-0.132***	-0.155***
	(0.022)	(0.032)	(0.033)	(0.032)	(0.043)
Age^2	0.000	0.001**	0.001*	0.001**	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Married	2.203***	2.718***	1.747***	2.718***	2.247***
	(0.018)	(0.039)	(0.038)	(0.039)	(0.038)
Young Children	-0.877***	-1.158***	-0.958***	-1.158***	-1.025***
	(0.020)	(0.036)	(0.036)	(0.036)	(0.036)
N° Children	0.098***	0.228***	0.222***	0.227***	0.189***
	(0.007)	(0.020)	(0.029)	(0.020)	(0.016)
Undergrad. Degree	0.361***	0.681***	0.517***	0.508***	0.286***
	(0.019)	(0.063)	(0.106)	(0.112)	(0.049)
Graduate Degree	0.440***	1.223***	1.486***	1.498***	0.402***
	(0.020)	(0.074)	(0.175)	(0.183)	(0.054)
White	-0.607***	-1.113***	0.000	-1.095***	-0.604***
	(0.016)	(0.062)	(.)	(0.062)	(0.043)
Mean(Undergrad.)				0.333**	
				(0.136)	
Mean(Graduate)				-0.347*	
				(0.199)	
Lagged(Employed)					1.538***
					(0.032)
$Employed_0$					2.534***
					(0.057)
σ_α		2.876		2.875	1.629
		(0.036)		(0.036)	(0.030)
ρ		0.715		0.715	0.446
		(0.005)		(0.005)	(0.009)
Observations	90,056	90,056	44,424	90,056	78,799
Log-Likelihood	-49,943.4	-36,142.1	-15,696.2	-36,134.0	-27,945.8

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors in parentheses.

1. Table 1 displays the results of different binary choice panel data models. Column 2 shows the results of the Static Random Effects Logit Model. We observe that all the variables in the model are statistically significant at least at the 5% level. Variables such as being married, the number of children, and holding a degree (either undergraduate or graduate level) have all positive coefficients, meaning that these variables increase the probability of being employed, with the married status being the most relevant among them. In contrast, `Age`, race (being white), or having young children decreases employment probability. Moreover, the effect of age seems to be non-monotonic, that is, the negative effect of aging is lower (less intense) for every additional age. To understand these effects' magnitudes, we would need to calculate marginal effects.

2. Column 1 of Table 1 reports the results of the Pooled Logit Model. We observe that each coefficient's signs are similar when compared to the RE Logit model, but with less marked magnitudes in this case. Under the standard RE assumptions -more specifically, Strict Exogeneity of the composite errors ε_{it}-, the RE Logit model is to be preferred when in presence of unobserved individual effects α_i, because it is essentially an FGLS estimator that takes unobserved heterogeneity into account to gain more efficiency. RE Model is a GLS version of Pooled OLS that accounts for the fact that composite errors are serially correlated due to the strong persistence of α_i across periods t. Consider the variance of β_{RE}:

$$V(\hat{\beta}_{RE}) = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' + \hat{\psi} T \sum_{i=1}^{N} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \right)^{-1}$$

with $\hat{\psi} = \frac{\hat{\sigma}_u^2}{T\hat{\sigma}_\alpha^2 + \hat{\sigma}_u^2}$. Here we see that the variance is an efficient weighted average of a between and within estimator and ψ tells us how important is between variation. Random effects will be always more efficient than FE and Pooled OLS if $\psi > 0$, resulting in smaller variance. On the other hand, if the standard random effects assumptions hold but the model does not actually contain individual effects α_i (in other words, $\psi = 1$), then V(β_{RE}) = V(β_{POLS}) and pooled OLS is efficient[1]. Finally, both models are inconsistent when Cov$(x_{it}, \alpha_i) \neq 0$, in which case the FE estimator is superior.

Breusch and Pagan (1980) developed a Lagrange Multiplier (LM) test to decide between random effects and Pooled OLS. The null hypothesis of this test is that the variance of the individual effects is zero. In this case, it is not possible to run the `xttest0` command, but in Table 1 we see that the σ_α^2 is statistically significant and it accounts for $\rho = 71.5\%$ of unobserved variance of y_{it}. Hence, there is evidence of significant differences across individuals, and the RE model is more likely a better model. If we still want to test this more formally, we can run a likelihood ratio test, defining the RE Logit as the more restrictive version of the Pooled Logit Model, such that:

$$H_0 : \text{Pooled Logit Model} \ \text{ vs. } \ H_1 : \text{Random Effects Model}$$

Using the log-likelihoods reported in Table 1, we calculate:

$$\text{LR} = 2[LL_{RE} - LL_{\text{Pooled}}] = 2[-36,142.1 + 49,943.4] = 26,602.6 > \chi^2_{1;0.95} = 3.84$$

As the statistic exceeds the critical $\chi^2_{1;0.95}$ value, we reject the null hypothesis and the RE Logit Model is preferred[2].

[1]Actually, the RE estimates here are less efficient than the pooled OLS estimates, which could be a sign that the RE model fails to model the error correlation well.

[2]The `xtlogit` command in STATA already performs an LR test of $\rho = 0$ by default, reaching the same conclusions as here.

3. The marginal effects of a logit regression are defined by:

$$\begin{aligned}\text{m.e.}(Age) &= \frac{\partial \Pr(y_{it}=1|x_{it},\alpha_i)}{\partial Age} = (\beta_{Age} + 2\beta_{Age^2} \times Age)\frac{\partial \Lambda(x_{it}'\beta + \alpha_i)}{\partial x_{it}'\beta} \\ &= (\beta_{Age} + 2\beta_{Age^2} \times Age)\Lambda(x_{it}'\beta + \alpha_i)[1 - \Lambda(x_{it}'\beta + \alpha_i)]\end{aligned}$$

Exploiting the fact that $\Lambda(x_{it}'\beta + \alpha_i) = 0.5$, we can calculate:

$$\begin{aligned}&(\beta_{Age} + 2\beta_{Age^2} \times Age)\Lambda(x_{it}'\beta + \alpha_i)[1 - \Lambda(x_{it}'\beta + \alpha_i)] \\ &= (-0.131 + 2 * (0.001) \times 40)(0.5)[1 - 0.5] = -0.01275 \approx\end{aligned}$$

In simple, the probability of employment at age 41 decreases by 1.275 percentage points, which means a final probability of $0.5 - 0.01275 = 0.48725$.

We can also calculate the point in age by which $\text{m.e.}(Age) = 0$:

$$\begin{aligned}(\beta_{Age} + 2\beta_{Age^2} \times Age)\Lambda(x_{it}'\beta + \alpha_i)[1 - \Lambda(x_{it}'\beta + \alpha_i)] &= 0 \\ \rightarrow \beta_{Age} + 2\beta_{Age^2} \times Age &= 0 \\ \rightarrow Age = -\frac{\beta_{Age}}{2\beta_{Age^2}} &= 65.5 \approx\end{aligned}$$

We see that at age 65.5, the marginal effects of age will be equal to 0.

4. Column 3 of Table 1 shows the results of the FE Logit Model. First, note that the variable `White` is excluded in the FE estimator since the time-demeaned transformation of the original data in any FE procedure eliminates everything that is time-invariant. Additionally, the FE estimator works with fewer observations, since the Conditional Maximum Likelihood drops individuals for whom y_{it} does not vary over time. Even though all coefficients are statistically significant, the effects of each variable are less pronounced than in the RE case, save for the `Graduate Degree` variable.

Normally, one should expect the RE to be an efficient estimator (lower standard errors)[3]. However, here we see that the standard errors for the `Married` variable are higher in the RE model than in the FE model. Therefore, it is not possible to perform a Hausman test to compare models, since its null hypothesis compares a consistent and efficient estimator (RE) against a consistent but less efficient estimator (FE), which is clearly not the case here. Alternatively, I re-run the models but exclude `Married` and `White` before performing a Hausman test in STATA (command `hausman`), whose results are:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = 6846.93^{***} > \chi^2_{5;0.95}$$

The test rejects H_0 of the difference between the coefficients not being systematic, thus, the FE estimator is the preferred model. The fact that the standard errors on the variable `Marriage` are lower in FE support in some degree this conclusion.

[3] Additionally, in terms of consistency, the random effects estimator $\hat{\beta}_{RE}$ converges to the within estimator $\hat{\beta}_{FE}$ as $T \rightarrow \infty$, since then $\theta \rightarrow 1$.

5. The results of the QFE Logit model are reported in Column 4 of Table 1. This QFE estimator is the Mundlak version, where the individual effects can be correlated with some x_{it} by adding their means over time:

$$\alpha_i = \overline{\text{undergrad}}_i'\delta_1 + \overline{\text{graduate}}_i'\delta_2 + \tilde{\alpha}_i$$

with $\tilde{\alpha}_i \sim N(0, \sigma^2_{\tilde{\alpha}})$ being the pure random effects. Mundlak assumes that δ_i (the effect of education) is the same in all periods t.

The results show coefficients highly similar to the ones from the RE Logit model, except for the education variables. The mean of having an undergraduate degree and a graduate degree are both significant, confirming the evidence of education being correlated with individual effects. The mean of `Graduate` has a negative effect on employment status in this case. Unobserved heterogeneity also appears to be the same as in the RE Logit model, explaining an important proportion of the unobserved variance. Finally, the efficiency is roughly the same compared to the RE model, except for the education variables where we observe worse precision (probably because of higher correlation induced by incorporating the mean education variables).

6. The Dynamic RE Logit Model using the Wooldridge approach treats the initial value y_{i1} as given but adds it to the model, by the assumption that:

$$\alpha_i = \lambda \text{Employed}_{i1} + \tilde{\alpha}_i$$

with $\tilde{\alpha}_i \sim N(0, \sigma^2_{\tilde{\alpha}})$ being the pure random effects. Column 5 in Table 1 reports the results of the Dynamic RE Model, confirming the relevance of state dependence as the coefficient from the lagged dependent variable is positive and significant (although the magnitude could be low). In that sense, employment in the previous period increases the probability of current employment.

Regarding unobserved heterogeneity, we need to construct or own ρ estimate:

$$\rho = \frac{\sigma^2_{\alpha}}{\sigma^2_{\alpha} + \sigma^2_u} = \frac{4.26}{4.26 + (\pi^2/3)} \approx 0.5642$$

where $V(a_i) = \lambda^2 V(y_{i1}) + \sigma^2_{\tilde{\alpha}} = (2.534)^2(0.493)(1-0.493) + (1.629)^2 \approx 4.26$. Because both $\sigma^2_{\tilde{\alpha}}$ and the coefficient associated with the initial value are significant, we can conclude that ρ also is. Here, unobserved heterogeneity explains around 56.42% of the unexplained variation in employment.

2 Tobit Model

1. Column 1 of Table 2 shows the results of the Static Random Effects Tobit Model. We use a censored model mainly because around 47.77% of the observations may be left-censored below 0. Generally, all the variables included report statistically significant results at the 1% level. Being female or white is associated with having less income, which makes economic sense in the first case, but the second case seems to be counterintuitive. On the other hand, age, education, and being married are strongly associated with higher income levels, with the `Married` variable being the most relevant here. Overall, these positive coefficients are also in line with intuition and economic literature. Finally, we observe again a non-monotonic incidence of age on earnings, that is, the positive effect of age on income decreases with age.

Table 2: Tobit Models Regression Results

Income	(1) RE Tobit 1	(2) RE Tobit 2	(3) RE Logit 3	(4) Dynamic Tobit
Sex	-16312.0***	-15847.4***	-8467.7***	-1883.0***
	(814.5)	(809.8)	(873.0)	(541.5)
Age	2222.9***	-145.5	2156.6***	1250.4***
	(341.9)	(495.6)	(340.6)	(452.4)
Age^2	-14.9***	-12.5**	-14.2***	-9.2*
	(4.3)	(5.8)	(4.3)	(5.6)
White	-11791.3***	-11376.9***	-11524.0***	-10454.6***
	(828.1)	(822.7)	(814.1)	(535.9)
Married	28065.6***	28116.5***	36649.4***	25868.9***
	(400.2)	(399.6)	(556.9)	(390.3)
Undergrad. Degree	10081.6***	10034.2***	9999.5***	3651.1***
	(749.8)	(747.7)	(741.8)	(585.9)
Graduate Degree	27468.3***	27546.3***	27360.1***	7126.4***
	(908.2)	(904.6)	(895.6)	(668.1)
Married × Sex			-17392.8***	
			(776.3)	
lagged(Income)				0.4***
				(0.0)
Income$_0$				1.0***
				(0.0)
Year Dummies		✓		
σ_α	39664.7***	39360.4***	38945.0***	22550.8***
	(368.0)	(365.3)	(363.6)	(276.6)
σ_u	29021.9***	28997.5***	28932.2***	28539.2***
	(103.8)	(103.7)	(103.5)	(110.8)
ρ	0.651	0.648	0.644	0.384
	(0.004)	(0.004)	(0.004)	(0.006)
Observations	88,750	88,750	88,750	75,005
Log-Likelihood	-571,539.0	-571,421.7	-571,289.3	-477,593.1

Notes: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors in parentheses. Year dummmies in Column (2) are added in the model, but not presented in the table.

2. The value of both the ρ statistic and σ_α are statistically significant, confirming the relevance of unobserved heterogeneity in income levels. In this case, we see that 65.1% of the unexplained variance in income is captured by individual effects α_i.

3. Column 2 in Table 2 shows the same model but adds year 7 dummies (the year 1994 is absorbed by the intercept) to check whether the model is better. All 7 dummies report positive and statistically significant coefficients. Additionally, I performed a Wald test after running the regression to check whether the year dummies are jointly significant, and it exceeds the $\chi^2_{7;0.95}$ critical value at 5%, confirming the relevance of the time dummies in the model. An LR test that compares Model 1 and 2 also rejects the null hypothesis, as 2[-571,421–571,539]$>\chi^2_{7;0.95}$. Now, adding these variables seems to have mainly changed the effect of age on income, no longer reporting statistically significant results. The remaining variables keep being significant and with the same signs as in Column 1, some of them with slightly less pronounced coefficients (save for marital status).

4. To study the potential heterogeneity in the effect of marital status on income, Column 3 reports the same RE Tobit model but incorporates the interaction between `Sex` and `Married`. In line with economic literature on gender gaps, there is a significant difference in the effect of marital status for males and females. From the coefficient of the interaction term, we see that married females are associated with decreased earnings, and from the `Married` variable only, we see that being married and male increases the level of earnings. Adding the interaction term leads to a slightly lower σ_α, but remains significant.

5. The Dynamic RE Tobit Model is presented in the last column of Table 2, based on the Wooldridge approach of treating the initial value y_{i1} as given but adding it into the model such that $\alpha_i = \lambda \text{Income}_{i1} + \tilde{\alpha}_i$, with $\tilde{\alpha}_i \sim N(0, \sigma^2_{\tilde{\alpha}})$ being the pure random effects. The results confirm the significant presence of state dependence, as the coefficient from the lagged Income is positive and significant (although the magnitude could be low in relative terms). Regarding unobserved heterogeneity, we need to construct or own ρ estimate. Using the same equation from question 1, we get:

$$\rho = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_u} = \frac{1.194e^9}{1.194e^9 + (29,229)^2} \approx 0.5164$$

where $V(a_i) = \lambda^2 V(y_{i1}) + \sigma^2_{\tilde{\alpha}} = (1.25)^2(18,522.9)^2 + (25,658.075)^2 \approx 1.194e^9$. Unobserved heterogeneity explains around 51.64% of the unexplained variation in employment. Because both σ_α and the coefficient associated with the initial value are significant, we can conclude that ρ also is.

6. To test whether the initial value is correlated with the individual effects α_i, we simply need to see if λ is statistically different from zero, which in this case is at the 1% level. A formal Wald test leads to the same conclusion. In this case, the initial value has a significant and positive effect on income.

3 Ordered Response Models

1. Column 1 of Table 3 reports the results of a Static Random Effects Ordered Probit Model, where the dependent variable is job satisfaction (1 =*'Like it very much'*, 2 =*'Like it fairly well'*, 3 =*'Dislike it somewhat'*, 4 =*'Dislike it very much'*). All variables are negative and statistically significant at the 1% level, except for the `Undergraduate Degree` variable, where we cannot reject the null hypothesis of zero effect. Income impacts negatively job dissatisfaction levels (hence the negative sign), as well as age and having a graduate degree.

2. Table 3 does not present an estimate for the constant term, mainly because the ordered response model needs to either fix the intercept to zero or fix the first cut-off m_1 to zero (normalization) in order for identification to be possible (Identification Problem). More generally, at least one parameter needs to be fixed for identification. STATA in this case chose to remove the intercept.

Table 3: Ordered Response Models Regression Results

	(1)	(2)	(3)	(4)
Job Satisfaction	RE Ologit	FE Ologit	RE Ologit	QFE Ologit
Age	-0.012***	-0.013***	-0.111***	-0.114***
	(0.002)	(0.003)	(0.029)	(0.029)
Income	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Undergrad. Degree	-0.071	-0.021	-0.058	-0.033
	(0.043)	(0.120)	(0.043)	(0.044)
Graduate Degree	-0.345***	-0.272	-0.312***	-0.240***
	(0.049)	(0.170)	(0.049)	(0.052)
Age^2			0.001***	0.001***
			(0.000)	(0.000)
Sex			-0.102**	-0.144***
			(0.040)	(0.041)
White			-0.044	-0.027
			(0.040)	(0.040)
Married			-0.148***	-0.138***
			(0.030)	(0.030)
Mean(Income)				-0.000***
				(0.000)
Cut 1	-0.712***		-2.792***	-2.949***
	(0.090)		(0.575)	(0.576)
Cut 2	2.650***		0.570	0.413
	(0.092)		(0.575)	(0.576)
Cut 3	4.210***		2.130***	1.974***
	(0.097)		(0.576)	(0.577)
σ^2_α	2.104***		2.087***	2.084***
	(0.058)		(0.058)	(0.058)
Observations	47,526	52,971	47,526	47,526
Log-Likelihood	-41047.4076	-20156.0992	-41024.1569	-41013.9830

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors in parentheses.

3. The Likelihood-Ratio Test performed by STATA and presented at the bottom of their output compares a Pooled Ordered Logit Model (H_0) against the Random Effects Ordered Logit Model (H_1) to determine which model is preferred. In particular, one can interpret this as a test of whether there is evidence of significant unobserved heterogeneity $\rho \neq 0$ or not. Here, the reported likelihood-ratio test follows a type of χ^2_1 distribution that exceeds the critical value[4] (`chibar2(01) = 7398.85`), showing that there is enough variability between individuals to favor a RE ordered Logit model over a Pooled RE Logit model.

4. Column 2 of Table 3 shows the results of the Fixed Effects Ordered Logit Model, which is a BUC estimator (Baetschmann, Staub & Winkelmann, 2015) and consists in the sum of the scores of different Chamberlain Conditional Maximum Likelihood (CML) estimators L_i^j. We observe that variables such as `Age` and `Income` remains significant relative to the RE model, and both now presents more intense magnitudes in their coefficients. `Undergraduate Degree` remains to be non-statistically significant, as in the RE model.

On the other hand, the conclusions regarding the variable `Graduate Degree` are different here, where now there is no evidence of a coefficient distinct from zero. This could be due to the fact that FE estimator eliminates everything time-invariant, and one can reasonably argue that education normally correlates with a number of unobserved individual effects. The $\hat{\beta}_{\text{Grad.}}$ in the RE estimator from Column 1 could have been absorbing some of this correlation, leading to inconsistency. Note also that all the standard errors are larger than the ones from the RE model, partly because of the BUC estimator uses a cluster-robust variance estimator that allows for correlation within the various contributions of individuals.

5. Recall that the statistic of the Hausman test is defined by:

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})'[V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE})$$

where $H \sim \chi^2_K$. With this we now calculate using the `Graduate Degree` variable (EDU_15):

$$H = \frac{(\hat{\beta}_{FE} - \hat{\beta}_{RE})^2}{V(\hat{\beta}_{FE}) - V(\hat{\beta}_{RE})} = \frac{(-0.272 - -0.345)^2}{0.170^2 - 0.049^2} \approx 0.201 < \chi^2_{1;0.95} = 3.841$$

The H_0 is not rejected because H does not exceeds the critical value, hence, there is no evidence that the FE model is preferred in this case.

6. Before choosing the best model, I propose two additional models that could be interesting to compare with. First, and to check whether we can improve the original RE model, I include additional variables to the first RE Ologit model such as `Age`^{2}, `Sex`, `White` and `Married` (see Column 3). Second, I run a Quasi-Fixed Effects Ologit Model using the Wooldridge approach (see Column 4), where I explained some of the individual effects as:

$$\alpha_i = \lambda \overline{\text{Income}}_i + \tilde{\alpha}_i$$

with $\tilde{\alpha}_i \sim N(0, \sigma^2_{\tilde{\alpha}})$ being the pure random effects. this Mundlak version assumes that the effect of income is the same in all periods t. It seems reasonable to say that some of the unobserved heterogeneity that affects job satisfaction can be explained by income differences across individuals.

First, we see that adding more controls leads to minor improvements in terms of log-likelihood, but enough to prefer it, as $2[-41,024.2 - -41,047.4] > \chi^2_{4;0.95} = 9.488$. Then, to test if the QFE

[4] For completeness, and according the the STATA guide about `chibar2`, the actual distribution of the LR test statistic is not the usual chi-squared with 1 degree of freedom but is instead a 50:50 mixture of a chi-squared with no degrees of freedom and a chi-squared with 1 degree of freedom.

Ordered Logit model represents a significant improvement compared to the initial RE Ordered Logit model, I performed a likelihood ratio test (with QFE Ologit chosen as an unconstrained model). The test result led me to reject the null hypothesis of equivalence between the models: $2[-41,014.0 - -41,047.4] > \chi^2_{1;0.95} = 3.841$.

Finally, to compare the QFE and FE Ordered models, we cannot perform an LR test, because they use different observations (nonnested). As a proxy, I can perform a Hausman test between Column (1) and Column (2). I do this with each variable:

- Age: $(-0.013 - -0.012)^2/(0.003^2 - 0.002^2) = 0.200 < 3.841$
- Undergraduate Degree: $(-0.021 - -0.071)^2/(0.120^2 - 0.043^2) = 0.199 < 3.841$
- Graduate Degree: $(-0.272 - -0.345)^2/(0.170^2 - 0.049^2) = 0.201 < 3.841$
- Income: $(-0.00000189 - -0.00000315)^2/(0.00000062^2 - 0.00000043^2) = \mathbf{7.958 > 3.841}$

We see that there is no evidence of the RE being a better model than the FE model, except regarding the variable `Income`. But the QFE model actually relaxes the RE assumption by allowing correlation between individual effects and income. When performing a Hausman test on this variable only (FE vs. QFE), we get that $(-0.00000189 + 0.00000198)^2/(0.00000062^2 - 0.00000051^2) \approx 0.0652$, and we cannot reject the null hypothesis of equivalence between these two models. My model of choice is therefore the QFE Ologit model.

7. The marginal effects of our ordered logit regression (using the QFE Ologit Model from Column 4) are defined by:

$$\begin{aligned}\frac{\partial \Pr(y_{it} = 1|x_{it}, \alpha_i)}{\partial \text{Graduate}} &= \frac{\partial}{\partial \text{Graduate}} \Pr(m_{j-1} < y^*_{it} < m_j|x_{it}, \alpha_i)\\ &= \frac{\partial}{\partial x'_{it}\beta} \Lambda(m_j - x'_{it}\beta - \alpha_i) - \Lambda(m_{j-1} - x'_{it}\beta - \alpha_i)\end{aligned}$$

$$= \beta_{\text{Grad.}}\{\Lambda(m_{j-1} - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_{j-1} - x'_{it}\beta - \alpha_i)] - \Lambda(m_j - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_j - x'_{it}\beta - \alpha_i)]\}$$

Exploiting the fact that $\Lambda(x'_{it}\beta + \alpha_i) = 0.5$, $m_0 = -\infty$ and $m_5 = \infty$, we can calculate:

$$\begin{aligned}&= \beta_{\text{Grad.}}\{\Lambda(m_0 - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_0 - x'_{it}\beta - \alpha_i)] - \Lambda(m_1 - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_1 - x'_{it}\beta - \alpha_i)]\}\\ &= (-0.240)\{-(0.5)[1 - 0.5]\} \approx 0.06\end{aligned}$$

In words, having a graduate degree *increases* by 6 percentage points the probability of being very satisfied with their job (*'Like it very much'*). In this case, it means a final probability of $0.5 + 0.06 = 0.56$.

$$\begin{aligned}&= \beta_{\text{Grad.}}\{\Lambda(m_4 - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_4 - x'_{it}\beta - \alpha_i)] - \Lambda(m_5 - x'_{it}\beta - \alpha_i)[1 - \Lambda(m_5 - x'_{it}\beta - \alpha_i)]\}\\ &= (-0.240)\{(0.5)[1 - 0.5]\} \approx -0.06\end{aligned}$$

Having a graduate degree *decreases* by 6 percentage points the probability of not being satisfied at all with their job (*'Dislike it very much'*). In this case, it means a final probability of $0.5 - 0.06 = 0.44$. These two cases basically represent opposite but symmetrical decisions.

STATA Code

```
/*****************************************************************************
ASSIGNMENT 2: PANEL DATA ANALYSIS OF MICROECONOMIC DECISIONS
DANIEL REDEL S.
DATE: Fall 2022
*****************************************************************************/
clear all

*ssc install blindschemes, replace all
*ssc install asdoc
*ssc install xtabond2
set scheme tab2, permanently

/*****************************************************************************
1. BINARY CHOICE MODELS
*****************************************************************************/

*----------------*
*--Load Data--*
*----------------*
clear
use " LOCAL DIRECTORY "

keep if WAVE <= 8 // My SNR = 2102630 ---> Waves 1 to 8
xtset ID WAVE

*------------------------------------------------------*
*--Question 1.1. Static Random Effects Logit Model --*
*------------------------------------------------------*
gen AGE2 = AGE^2
global covariates AGE AGE2 MARRIED YOUNG_CH NUM_CH EDU_13_15 EDU_15 WHITE

xtlogit EMPL $covariates, re
est store re_logit

*-------------------------------------*
*--Question 1.2. Pooled Logit Model --*
*-------------------------------------*
logit EMPL $covariates
est store pooled_logit

*-----------------------------------------------*
*--Question 1.2.a. Pooled vs Random Effects --*
*-----------------------------------------------*
* See Document
```

```
*——————————————————————————*
*——Question 1.3. Marginal Effects ——*
*——————————————————————————*
dis (-0.131+2*(0.001)*40)*(0.5)*[1-0.5]

*————————————————————————————————————*
*——Question 1.4. Static Fixed Effects Logit Model——*
*————————————————————————————————————*
xtlogit EMPL $covariates, fe
est store fe_logit
ereturn list

*——————————————————————————*
*——Question 1.4.a. HAUSMAN Test ——*
*——————————————————————————*
xtlogit EMPL AGE AGE2 YOUNG_CH NUM_CH EDU_13_15 EDU_15, re
est store re_haus
xtlogit EMPL AGE AGE2 YOUNG_CH NUM_CH EDU_13_15 EDU_15, fe
est store fe_haus

hausman fe_haus re_haus

*————————————————————————————————————*
*——Question 1.5. Quasi-Fixed Effects Logit Model——*
*————————————————————————————————————*
** Mundlak Version
egen MEAN_EDU_13_15 = mean(EDU_13_15), by(ID)
egen MEAN_EDU_15 = mean(EDU_15), by(ID)

xtlogit EMPL $covariates MEAN_EDU_13_15 MEAN_EDU_15, re
est store qfe_logit

*————————————————————————————————————————————————*
*——Question 1.6. Dynamic Random Effects Logit Model (Wooldridge) ——*
*————————————————————————————————————————————————*
* Lagged Value *
gen lagged_EMPL=.
replace lagged_EMPL=l.EMPL if WAVE>1

* Initial Value *
gen EMPL_0 = EMPL
by ID, sort: replace EMPL_0 = EMPL[1]

* Model *
xtlogit EMPL $covariates lagged_EMPL EMPL_0 if WAVE > 1, re
* xtlogit EMPL $covariates lagged_EMPL EMPL_0, re // should be the same
est store dynamic_re_logit
```

```
*-------------------------------------------*
*--Question 1.6.a. State Dependence --*
*-------------------------------------------*
* See Document

*----------------------------------------------------*
*--Question 1.6.b. Unobserved Heterogeneity --*
*----------------------------------------------------*
sum EMPL if WAVE==1
sum EMPL_0 // V(y) == 0.493

dis (2.534)^2*(0.493)*(1-0.493)+(1.629)^2 // New Sigma2_alpha
dis (4.2586154/(4.2586154+(3.14159^2/3))) // New Rho

*----------------------------------------------------------*
*--Question 1. FINAL TABLE: BINARY CHOICE MODELS --*
*----------------------------------------------------------*
esttab pooled_logit re_logit fe_logit qfe_logit dynamic_re_logit ///
        using table1.tex, cells(b(star fmt(%9.3f)) se(par fmt(%9.3f)))
///
        mtitles("Pooled Logit" "RE Logit" "FE Logit" "QFE Logit" "Dynamic") ///
        nonumbers starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) ///
        stats(N ll sigma_u rho, fmt(a4 4 a4) ///
        labels("Observations" "Log-Likelihood" "sigma_u" "rho")) replace

/**************************************************************************
2. TOBIT MODEL
**************************************************************************/
est clear

**See Censored
gen censored = 0
replace censored = 1 if INCOME==0
replace censored = 0 if INCOME == .
tab censored // 46.77 are censored
tab EMPL // 45.39 unemployed

*------------------------------------------------------------*
*--Question 2.1. Static Random Effects Tobit Model --*
*------------------------------------------------------------*
global covariates_tobit SEX AGE AGE2 WHITE MARRIED EDU_13_15 EDU_15
xttobit INCOME $covariates_tobit, ll(0)
est store tobit1

*-------------------------------------------------*
*--Question 2.2. Unobserved Heterogeneity --*
*-------------------------------------------------*
* See Document
```

```
*---------------------------------------------------------------------------*
*--Question 2.3. Static Random Effects Tobit Model: Year Dummies --*
*---------------------------------------------------------------------------*
xttobit INCOME $covariates_tobit i.YEAR, ll(0)
est store tobit2

*Dummy for each Year
forvalues i = 1(1)8{
 generate year`i' = 0
 replace year`i' = 1 if YEAR == 1992+2*`i'
 }

* RE Tobit
xttobit INCOME $covariates_tobit ///
year2 year3 year4 year5 year6 year7 year8, ll(0)
est store tobit2

*Test whether they are jointly significant
test year2 year3 year4 year5 year6 year7 year8
/* Time dummies are jointly significantly different from zero (at the 5%)*/

*----------------------------------------------------------------------------*
*--Question 2.4. Static Random Effects Tobit Model: Marital Status --*
*----------------------------------------------------------------------------*
gen MARRIED_SEX = MARRIED*SEX
xttobit INCOME $covariates_tobit MARRIED_SEX, ll(0)
est store tobit3

*-----------------------------------------------------------*
*--Question 2.5. Dynamic Random Effects Tobit Model --*
*-----------------------------------------------------------*
* Lagged Value *
gen lagged_INCOME=.
replace lagged_INCOME=l.INCOME if WAVE>1

* Initival Value *
gen INCOME_0 = INCOME
by ID, sort: replace INCOME_0 = INCOME[1]

* Model *
xttobit INCOME $covariates_tobit lagged_INCOME INCOME_0 if WAVE > 1, ll(0)
est store tobit4
return list
matrix list r(table)
matrix list e(b)

*----------------------------------------------------*
```

```
*--Question 2.5.1. Unobserved Heterogeneity --*
*-----------------------------------------------*
sum INCOME_0 // V(y_0) = 18522.89

* New Sigma Alpha
dis _b[INCOME_0]^2*(18522.89)^2 + _b[/: sigma_u]^2
* New RHO:
dis "RHO: "(_b[INCOME_0]^2*(18522.89)^2 ///
        + _b[/: sigma_u]^2)/(_b[INCOME_0]^2*(18522.89)^2 ///
        + _b[/: sigma_u]^2 + _b[/:sigma_e]^2)

*--------------------------------------------*
*--Question 2. FINAL TABLE: TOBIT MODELS --*
*--------------------------------------------*
esttab tobit1 tobit2 tobit3 tobit4 using table2.tex ///
        , cells(b(star) se(par fmt(%9.3f))) ///
        mtitles("RE Tobit 1" "RE Tobit 2" "RE Logit 3" "Dynamic T") nonumbers ///
        starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) modelwidth(12) ///
        stats(N ll sigma_u rho, fmt(a4 4 a4) ///
        labels("Observations" "Log-Likelihood" "sigma_u" "rho")) replace

*------------------------------------------------------------------*
*--Question 2.6. Individual Effects against Initial Value --*
*------------------------------------------------------------------*
test INCOME_0
/* Time dummies are jointly significantly different from zero (at the 5%)*/

/**************************************************************************
2. ORDERED RESPONSE MODELS
**************************************************************************/

*-----------------------------------------------------------*
*--Question 3.1. Static Random Effects Ordered Logit --*
*-----------------------------------------------------------*
xtologit JOB_SAT AGE INCOME EDU_13_15 EDU_15 if EMPL==1
est store re_ologit

*-----------------------------*
*--Question 3.2. Constant Term --*
*-----------------------------*
* See Document

*--------------------------------------------------------*
*--Question 3.3. Likelihood Ratio Test in Output --*
*--------------------------------------------------------*
* See Document

*--------------------------------------------------------*
```

```
*--Question 3.4. Static Fixed Effects Ordered Logit --*
*------------------------------------------------------------*
feologit JOB_SAT AGE INCOME EDU_13_15 EDU_15 if EMPL==1
est store fe_ologit

*------------------------------------------------------------*
*--Question 3.5. Hausmann Test RE vs FE: Education --*
*------------------------------------------------------------*
dis (-0.272--0.345)^2/(0.170^2-0.049^2) // 0.201

*------------------------------------------------------*
*--Question 3.6.1. RE ORDERED RESPONSE MODEL 2 --*
*------------------------------------------------------*
xtologit JOB_SAT AGE AGE2 INCOME SEX WHITE MARRIED EDU_13_15 EDU_15 if EMPL==1
est store re_ologit1

*------------------------------------------------------*
*--Question 3.6.2. QFE ORDERED RESPONSE MODEL --*
*------------------------------------------------------*
*** Quasi-Fixed Effects: Mundlak Version
egen MEAN_INCOME = mean(INCOME), by(ID)

xtologit JOB_SAT AGE AGE2 INCOME SEX WHITE MARRIED ///
        EDU_13_15 EDU_15 MEAN_INCOME if EMPL==1
est store qfe_ologit

*-------------------------------------------*
*--Question 3.6.3. MODEL ASSESSMENT --*
*-------------------------------------------*
 ** RE1 Model vs RE2 Model, 4 degrees-of-freedom
 dis 2*(-41024.1569--41047.4076)

** RE2 Model vs QFE Model, 1 degree-of-freedom
 dis 2*(-41013.9830--41047.4076)

** FE Model vs RE Model
dis (-0.013--0.012)^2/(0.003^2-0.002^2)
dis (-0.021--0.071)^2/(0.120^2-0.043^2)
dis (-0.272--0.345)^2/(0.170^2-0.049^2)
** Income:
dis (-0.00000189--0.00000315)^2/(0.00000062^2-0.00000043^2)

** FE Model vs QFE Model
** Income :
dis (-0.00000189--0.00000198)^2/(0.00000062^2-0.00000051^2) // == 0.065

**format fmt
esttab fe_ologit re_ologit qfe_ologit ///
```

```
        , cells(b(star fmt(%9.8f)) se(par fmt(%9.8f))) ///
        mtitles("FE Ologit" "RE Ologit") nonumbers ///
        starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) ///
        modelwidth(12) stats(N ll sigma_u, fmt(a4 4 a4) ///
        labels("Observations" "Log-Likelihood" "sigma_u"))

*------------------------------------------------------*
*--Question 3.7.  Marginal Effects of Education --*
*------------------------------------------------------*
*Using QFE Ologit
dis -(0.5)*(1-0.5)*(-0.240)
dis  (0.5)*(1-0.5)*(-0.240)

*---------------------------------------------------------*
*--Question 3. FINAL TABLE: ORDERED RESPONSE MODELS --*
*---------------------------------------------------------*
esttab re_ologit fe_ologit re_ologit1 qfe_ologit using table3.tex///
        , cells(b(star fmt(%9.3f)) se(par fmt(%9.3f)))  ///
        mtitles("RE Ologit 1" "FE Ologit 1" "RE Ologit 2" "QFE Ologit") ///
        starlevels(* 0.1 ** 0.05 *** 0.01) varwidth(16) modelwidth(12) ///
        stats(N ll sigma_u, fmt(a4 4 a4) ///
        labels("Observations" "Log-Likelihood" "sigma_u")) replace
```