Panel Data Analysis for Microeconomic Decision

Mattia Zen – m.zen@tilburguniversity.edu – Snr: 2132868

Assignmenet 2

---

1 Question 1

1.1

To clean the data set I used these lines of code:

```
egen hh = group(ident)
drop if hh<881
drop if hh>2098
drop hh
```

1.2

Wave	dstock=1	total observations	**Ownership ratio**
1	610	1,218	**0.501**
2	617	1,218	**0.507**
3	589	1,218	**0.484**
4	599	1,218	**0.492**
5	576	1,218	**0.473**
Total	2991	6,090	**0.491**

1.3

We conclude that the majority of households do not hold any shares. However interesting result is that the top 25% invest at least half of their financial wealth.

Wave	25%	50%	75%
1	0	0.0003	0.476
2	0	.0116	0.567
3	0	0	0.519
4	0	0	0.593
5	0	0	0.5

	re1 b	re2 b	re3 b	re4 b
dstock				
female	-.0704329	-.0477283	-.0483862	-.1096444
age	-.0235671	-.0121673	-.0119012	-.0428574
edyrs	.110408	.1122267	.1122479	.1966052
nonwhite	-.2533199	-.2470472	-.2473678	-.4471121
hispanic	-.9153919	-.8960443	-.8964363	-1.672613
marrlt	.1722675	.1789723	.179057	.3245703
ltotfw	.6533333	.654891	.6534943	1.174973
1.wave		0		
2.wave		.0851119		
3.wave		-.0404822		
4.wave		-.0931376		
5.wave		-.1053157		
wave			-.0387979	
_cons	-7.13671	-7.866015	-7.782259	-12.80559
/				
lnsig2u	.6126751	.6142398	.6125573	1.746525

Figure 1: Coefficient for all the regression run to answer question 2

2 Question 2

2.1

This table shows which part of the population is more likely to have any shares. It is interesting to highlight that the result for the variables `female`, `nonwhite`, `hispanic` and `marrlt` are not statistically significant at a 95% confidence interval. In my opinion, the most interesting result is the statistically significant negative correlation with age. This means that young people are more likely to own any stock.

2.2

To understand the importance of unobserved heterogeneity we need to look at ρ.

dstock	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
female	-.0704329	.1017347	-0.69	0.489	-.2698292	.1289634
age	-.0235671	.0065113	-3.62	0.000	-.0363289	-.0108053
edyrs	.110408	.0202869	5.44	0.000	.0706463	.1501696
nonwhite	-.2533199	.1831514	-1.38	0.167	-.6122901	.1056504
hispanic	-.9153919	.3596212	-2.55	0.011	-1.620236	-.2105473
marrlt	.1722675	.10204	1.69	0.091	-.0277272	.3722621
ltotfw	.6533333	.0250839	26.05	0.000	.6041698	.7024968
_cons	-7.13671	.5551752	-12.85	0.000	-8.224833	-6.048586
/lnsig2u	.6126751	.0863066			.4435173	.7818329
sigma_u	1.358441	.0586212			1.24827	1.478335
rho	.6485508	.0196721			.6090968	.686075

LR test of rho=0: chibar2(01) = 1085.69 Prob >= chibar2 = 0.000

Figure 2: Estimation for the answer in question 2.1

$$\rho = \frac{\hat{\sigma}_{\alpha}^2}{\hat{\sigma}_{\alpha}^2 + \hat{\sigma}_{\epsilon}^2} \tag{1}$$

In this case, a value of 0.648 means that roughly 65% of the total variation is due to unobserved individual-specific effect.

2.3

Once we estimate the parameter per each wage (you can see the results above) to test if they are jointly significant we can compute the likelihood ratio between the restricted and the unrestricted model. To achieve this we used the `lrtest` command in Stata. This results in a p-value of 0.2556, which means that we cannot reject that the two models are statistically different from each other.

2.4

To choose between the three models, we can calculate different Likelihood Ratios for each model and compare them two by two. During none of the comparisons, we achieve statistically significant results. this suggests that the three models do not differ from each other.

2.5

Between logit and probit, I choose the model that best fits my data, thus I will look at the goodness of fit. In this case, I just chose the model with the highest Log-Likelihood, which in this case is the logit one (-2590 vs -2596).

3 Question 3

3.1

```
gen share_0 =share
by ident, sort: replace share_0 =share[1]
gen dstock_0 = (share_0>0)

xtset ident wave
xtprobit dstock L.dstock female age edyrs nonwhite hispanic marrlt ltotfw if ident>1, re
```

dstock	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
dstock						
L1.	1.324252	.0668893	19.80	0.000	1.193151	1.455352
female	.0283282	.0565561	0.50	0.616	-.0825198	.1391762
age	-.0159784	.005039	-3.17	0.002	-.0258545	-.0061022
edyrs	.0313279	.0114891	2.73	0.006	.0088097	.0538461
nonwhite	.0169266	.1014914	0.17	0.868	-.1819929	.2158462
hispanic	-.3647097	.2066101	-1.77	0.078	-.7696581	.0402387
marrlt	.1033843	.0644747	1.60	0.109	-.0229838	.2297525
ltotfw	.4530015	.0284188	15.94	0.000	.3973017	.5087013
_cons	-5.061662	.4379962	-11.56	0.000	-5.920119	-4.203205
/lnsig2u	-1.925791	.4966888			-2.899283	-.9522988
sigma_u	.3817858	.0948144			.2346544	.6211707
rho	.1272172	.0551488			.052189	.2784227

LR test of rho=0: chibar2(01) = 5.63 Prob >= chibar2 = 0.009

Figure 3: Stata output from the regression run to answer 3.1

3.2

As discussed before we can use equation 1 to understand the importance of individual effect in the model. In this case, we have $\rho = 0.127$, which is significantly smaller than the result we found above. The reason for this difference is the inclusion of the lagged dependent variable in the model. This variable accounts for the consistency of y_i over time and it captures part of the variation previously attributed to α_i

3.3

The coefficient associated with the lagged dependent variable, `L-dstock` is statistically significant (p-value=0.000). This test for the presence of the state dependence, meaning that owning a stock in period $t-1$ influence (in this case increase since γ is positive) the likelihood of owning a stock in time t.

3.4

Recalling from the first question the ownership rate for the entire sample is 0.491.

$$p(dstock_{it} = 1|x_{it}, \alpha_i, y_{it-1}) = 0.491$$

We also know that to compute the marginal effect we need γ which is 1.324.

$$ME = \gamma * p(dstock_{it} = 1|x_{it}, \alpha_i, y_{it-1}) * [1 - p(dstock_{it} = 1|x_{it}, \alpha_i, y_{it-1})]$$
$$= 1.324(0.491)(1 - 0.491) = 0.331$$

3.5

4 Question 4

	tobit1 b/se	tobit2 b/se	tobit3 b/se
share			
female	-.0231798	-.0241877	-.0256121
	(.0274575)	(.0276454)	(.0279422)
age	-.0039321*	-.0034799*	-.0035795*
	(.001674)	(.0016833)	(.0017075)
edyrs	.0283604***	.0295054***	.0297854***
	(.0054408)	(.0054872)	(.0055462)
nonwhite	-.065232	-.0639621	-.0637015
	(.0507317)	(.0512172)	(.0517614)
hispanic	-.2526776*	-.2435347*	-.2468032*
	(.1011017)	(.1020668)	(.1031447)
marrlt	.0409921	.0387789	.0399167
	(.0270019)	(.0271637)	(.0275052)
ltotfw	.1712154***	.3801983***	.3823756***
	(.0059161)	(.0506934)	(.0513181)
ltotfw_sq		-.0097084***	-.009781***
		(.002324)	(.0023532)
_cons	-1.992548***	-3.134958***	-3.149802***
	(.1421692)	(.3116464)	(.3154314)
/			
sigma_u	.3841911***	.3869662***	.3905432***
	(.0124674)	(.0125897)	(.0127491)
sigma_e	.371699***	.371334***	.3771156***
	(.0056028)	(.0056023)	(.0057558)

Figure 4: Table showing the results of the three model estimated in Question 4

4.1

We note from the result shown above that few of the coefficients are statistically significant. Comparing to the results mentioned in section 2.1 in this model the variable `hispanic` is statistically significant at a 95% CI. The negative coefficient highlight e lower probability for this part of the population of earning stocks.
It is also interesting to note the difference in the magnitude for the variable `ltotfw`, which in the previous model was over 0.65, while in the Tobit model is roughly 0.17.

espandi sul perché

4.2

Starting from equation 1 we can derive the inverse equation:

$$\hat{\sigma}_\alpha^2 = \frac{\hat{\sigma}_\epsilon^2}{1-\rho}$$

Since from the state output we have $\rho = 0.517$ and $\hat{\sigma}_\epsilon^2 = 0.371^2$ we can substitute those values in the above mentioned formula and find the value of $\hat{\sigma}_\alpha^2$ which is 0.285.
In this case, the unobserved heterogeneity accounts for over 50% of the total variation. This is roughly 15% less than what we found in the same regression for the probit model.
To compute the marginal effect of `totfw` with 1% difference we start with $p(y_{it}^* > 0|x_{it}, \alpha_i) = 0.5$.

$$ME = \beta_{ltotfw} * p(share^* > 0|x_{it}, \alpha_i)$$
$$ME = 0.171 * 0.5 = 0.0855$$

Since the variable `ltotfw` enters the model in logarithmic form the marginal effect is expressed in % terms. This means that an increase in total financial wealth by 1% increases the probability of owning stocks by 0.0855.

4.3

```
gen ltotfw_sq=ltotfw^2
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq, ll(0)
est store tobit2

lrtest tobit1 tobit2
```

According to the likelihood ratio computed the difference between the two models is statistically significant with a p-value of 0.0000. This means that we can reject the null hypothesis of the two models being equal.

4.4

4.5

```
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq, ll(0) ul(1)
```

The results of the model are very similar to the one previously developed.
This result is not surprising since the Tobit model censored from 0 to 1 perfectly fits the distribution of share. Since the variable `share` represents the probability of owning a share its natural upper limit is 1. Thus even without the upper censoring, the model estimates similar coefficients since the maximum value that the latent variable, y^*_{it} reaches is 1.

$$y^*_{it} = x'_{it}\beta + \alpha_i + \epsilon_{it}$$

5 Question 5

5.1

To account for correlation between the individual effect and the variable `ltotfw` we can use a Quasi Fixed effect approach. In this case we model the individual effect α_i, with one random part, $\tilde{\alpha}_i$, and a component correlated with the variable, $\delta\bar{x}'_i$.

$$\alpha_i = \tilde{\alpha}_i + \delta\bar{x}'_i$$

This model is programmed in Stata with the code written below:

```
egen ltotfw_mean = mean(ltotfw), by(ident)
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq ltotfw_mean,
ll(0) ul(1)
```

share	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
female	-.0201522	.0280035	-0.72	0.472	-.075038	.0347337
age	-.0036674	.0017086	-2.15	0.032	-.0070162	-.0003186
edyrs	.0212872	.0057599	3.70	0.000	.0099979	.0325764
nonwhite	-.017443	.0526053	-0.33	0.740	-.1205476	.0856615
hispanic	-.2058118	.1039758	-1.98	0.048	-.4096007	-.002023
marrlt	.0294693	.0275985	1.07	0.286	-.0246228	.0835614
ltotfw	.3812654	.0513476	7.43	0.000	.2806259	.4819049
ltotfw_sq	-.0106142	.0023607	-4.50	0.000	-.0152411	-.0059874
ltotfw_mean	.0623083	.0117821	5.29	0.000	.0392159	.0854007
_cons	-3.587283	.3287677	-10.91	0.000	-4.231655	-2.94291
/sigma_u	.3899326	.0127693	30.54	0.000	.3649053	.4149599
/sigma_e	.3769113	.0057492	65.56	0.000	.3656431	.3881795
rho	.5169755	.0175564			.4825451	.5512797

LR test of sigma_u=0: chibar2(01) = 1203.07 Prob >= chibar2 = 0.000

Figure 5: Result from the Regression for Question 5

5.2

To test for the statistical significance of the `ltotfw_mean` we can look at the p-value of the coefficient in the Stata output. Since the inclusion of the variable is statistically significant ($p < 0.05$) we can state that there is an individual-specific effect of wealth.

5.3

In the last regression, we estimated a coefficient of 0.381, while in the first regression of the previous exercise, we estimated a coefficient of 0.171. This difference can be due to the increase in the number of variables which separate the effect of `ltotfw` with its quadratic form and the individual specific effect of it. However, we note that the difference is small if we compare this result with the last regression run in the previous exercise. Looking more in detail the comparison between these two models we note that adding the individual-specific effect of financial wealth in the regression decreases the coefficient of other variables (such as `nonwhite`) suggesting that those variables captured the effect.

5.4

In this case the we cannot use ρ as before since it would estimate the fraction of $\sigma_{\tilde{\alpha}}$, instead of σ_α. To achieve this we firstly need to compute σ_α as $var(\delta\bar{x}_i + \sigma_{\tilde{\alpha}}$.

$$\hat{\sigma}_\alpha^2 = \hat{\delta}^2 var(\bar{x}_i) + \hat{\sigma}_{\tilde{\alpha}}^2 = 0.062^2 * 0.011^2 + 0.389^2 = 0.1513$$

From this value, I can calculate the share of the variance of the individual effect using ρ.

$$\rho = \frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\epsilon^2 + \hat{\sigma}_\alpha^2} = \frac{0.1513}{0.3769 + 0.1513} = 0.2864$$

Stata Code

```
****************************************
*             ASSIGNEMENT 2
*             Mattia Zen
****************************************
clear all
capture log close

log using zen_assignment2.log, text replace

************
* Question 1
************

* Importing the dataset

cd "C:\Users\matti\OneDrive\Desktop\uni\MSc\Panel Data\Assignements\Assignement2"
use stocks_balanced-1

* 1.a

egen hh = group(ident)
drop if hh<881
drop if hh>2098
drop hh

xtset ident wave

* 1.b
gen dstock = (share>0)
by wave, sort : count if dstock==1



* 1.c
by wave, sort: summarize share, detail
* most of the people not own any shares

************
* Question 2
************

* 2.a
gen ltotfw =ln(totfw)

xtprobit dstock female age edyrs nonwhite hispanic marrlt ltotfw, re
```

```
est store re1
esttab using 2_a.tex, replace


* 2.b


* 2.c
xtprobit dstock female age edyrs nonwhite hispanic marrlt ltotfw i.wave, re
est store re2
esttab using 2_c.tex, replace

lrtest re2 re1

* 2.d
xtprobit dstock female age edyrs nonwhite hispanic marrlt ltotfw wave, re
est store re3

* LR test
lrtest re1 re2
lrtest re3 re2
lrtest re1 re3


* 2.e
xtlogit dstock female age edyrs nonwhite hispanic marrlt ltotfw, re
est store re4


estout re1 re2 re3 re4
esttab re1 re2 re3 re4 using 2.tex, r2 replace


************
* Question 3
************
*3.a
gen share_0 =share
by ident, sort: replace share_0 =share[1]

gen dstock_0 = (share_0>0)

xtset ident wave

xtprobit dstock L.dstock female age edyrs nonwhite hispanic marrlt ltotfw if ident>1, re
esttab using 3a.tex, replace
```

```
*3.e
xtprobit dstock L.dstock female age edyrs nonwhite hispanic marrlt ltotfw if ident>1, re



************
* Question 4
************

* 4.a
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw, ll(0)
est store tobit1


* 4.c
gen ltotfw_sq=ltotfw^2
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq, ll(0)
est store tobit2

lrtest tobit1 tobit2

* 4.d
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq, ll(0)
margins, dydx(ltotfw) at(ltotfw=(min(max)))
marginsplot

* 4.e
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq, ll(0) ul(1)
est store tobit3

estout tobit1 tobit2 tobit3, cells(b(star) se(par))

************
* Question 5
************
* 5.a
egen ltotfw_mean = mean(ltotfw), by(ident)
xttobit share female age edyrs nonwhite hispanic marrlt ltotfw ltotfw_sq ltotfw_mean, ll(0) ul(1)
est store tobit4

lrtest tobit4 tobit1
```