Homework - Week 7

Victor Jiao – `bjiao@uchicago.edu`

November 26, 2024

This assignment is not in the format of ”answer a list of questions” as is in previous assignments. I will divide the sections in my own way to allow a comprehensive understanding of this data task. These sections consist of explanations of the Do file structure, functions I have used, results, as well as discoveries/conclusions derived from results (if requested).

1 README

This section provides all relevant information to the Do file, titled `homework-victor.do`. You may find the Do file in the homework folder after extraction of the zip file. After opening the Do file, you may navigate to different sections using bookmarks.

The setup section sets working directories, creates a log file, installs `esttab` package, configures Stata versions, adds colors to commands, and most importantly, reads in `dohmen.dta`. Note that if you want to run the Do file on your own device, you *must* change the working directory to inside your homework folder, wherever it is at on your device. I can't find ways to get around that. Nevertheless, all file paths are relative in the remaining portions of this file, so the working directory is the only thing you have to change for everything to work.

The descriptive functions section, as the name suggests, contains functions that describe variables used in the task. They need not be run.

The tasks sections below run relevant interval regressions, as requested by the assignment. I will describe in subsequent sections in greater detail.

2 Replication of Table 2 of Dohmen et al. (2011)

This section provides results from Task 1 of the assignment.

I used three variables in the dataset for this regression model: `row_lo`, the lower row number of the switch; `row_hi`, the upper row number of the switch; and `risk_attitude`, the survey risk attitude. The first two variables create a dependent interval that captures the upper and lower bounds of the true value of the stable payout that one decides to make the switch. `risk_attitude` is the independent covariate used.

We can write the regression in mathematical terms:

$$\texttt{row} = \beta_0 + \beta_1 * \texttt{risk\_attitude}$$

`row` represents the [`row_lo, row_hi`] interval. As the assignment says, this regression is essentially an OLS regression, but the dependent is an interval.

We generate the regression in Stata by running

```
intreg row_lo row_hi risk_attitude
```

and we can generate the exact result as in the paper. The result is below:

Table 1: Interval regression between row switch and risk attitude, homoskedastic

	Lower row number of the switch
model	
Willingness to Take Risks (0-10)	0.611^{***}
	(0.123)
Constant	5.919^{***}
	(0.661)
lnsigma	
Constant	1.867^{***}
	(0.0374)
Log-Likelihood	-1348.3
Observations	450

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The estimation results demonstrate a positive and statistically significant relationship between survey risk attitude and row where individuals switch from choosing a lottery to opting for a stable income in the incentivized experiment (captured by [`row_lo` and `row_hi`]). Specifically, the coefficient of 0.611 on (`risk_attitude`) means that for each additional unit increase in the risk attitude score, the expected switch point to the stable income increases by approximately 0.611 units. The findings corroborate with that of the Dohmen paper. In terms of validation, these findings support the claim that the hypothetical survey instrument is a valid measure of risk attitudes. The survey scores align with actual decision-making behavior observed in the incentivized experiment. In other words, participants' self-reported risk attitudes are effective predictors to their willingness to take risks.

3 The Heteroskedastic Option

This section provides results from Task 2 of the assignment.

As suggested by the assignment, I estimated a full model that allows for heteroskedasticity, by including a heteroskedastic option in the code:

```
intreg row_lo row_hi risk_attitude het(crra_mid)
```

My understanding is that heteroskedastic option cannot accommodate intervals, so I have to use point estimate `crra_mid` instead of `[crra_lo, crra_hi]`. The dependent variable is still an interval, so hopefully I don't fail the assignment. I believe that this setup relaxes the homoskedasticity assumption, and can allow survey risk attitude to affect CRRA estimations.

The result is below:

Table 2: Interval regression between row switch and risk attitude, heteroskedastic

	Lower row number of the switch
model	
Willingness to Take Risks (0-10)	0.189^{***}
	(0.0719)
Constant	2.676^{***}
	(0.354)
lnsigma	
Constant	2.624^{***}
	(0.0894)
Log-Likelihood	-1252.2
Observations	450

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The positive coefficient for `risk_attitude` corroborates with the previous model that individuals with higher self-reported risk tolerance tend to switch from the lottery to the stable income at higher stable payouts. Specifically, for each additional unit increase in the risk attitude score (which ranges from 0 to 10), the expected switching point increases by approximately 0.19 rows.

The negative and highly significant coefficient for `crra_mid` in the variance equation implies that as the estimated CRRA increases, the variance of the error term decreases. In other words, individuals with higher levels of risk aversion not only switch earlier to the stable income but also exhibit less variability in their switching behavior. Their decisions are more consistent compared to those with lower risk aversion. On the flip side, more risk-tolerant individuals show greater variability, possibly reflecting more diverse risk-taking strategies.

4 Estimation with CRRA Intervals

This section provides results from Task 3 of the assignment. My understanding is that I need to repeat the estimation from Task 1, but using CRRA intervals as dependent variable. I'm not sure if I can include a heteroskedasticity condition for CRRA when CRRA is the dependent variable. As a result, I should not consider repeating the estimation from Task 2. And as the assignment suggested, "we see if the survey response is related to a measure of risk preferences under EUT," I believe I can comfortably assume homoskedasticity. The code is:

```
intreg crra_lo crra_hi risk_atittude
```

Likewise, I've also stored the model as `crra_homoskedastic_model`.

The result is presented below:

Table 3: Interval regression between CRRA estimations and risk attitude, homoskedastic

	Lower bound of CRRA
model	
Willingness to Take Risks (0-10)	-0.0377***
	(0.00814)
Constant	0.604***
	(0.0439)
lnsigma	
Constant	-0.845***
	(0.0375)
Log-Likelihood	-1390.4
Observations	450

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The coefficient for `risk_attitude` is -0.0377 with a standard error of 0.00814, and is highly significant. This indicates that for each additional unit increase in the self-reported risk tolerance, the estimated CRRA decreases by approximately 0.0377.

These findings imply that individuals who report higher risk tolerance on the survey indeed exhibit lower levels of risk aversion in the incentivized experiment, as reflected by the decreased CRRA estimations. The significant negative coefficient shows that self-reported risk attitudes align closely with actual risk-taking behavior observed experimentally. This alignment confirms the participants' self-reported risk attitudes in capturing true risk preferences.

I have also included how the predicted mean of CRRA varies with the hypothetical survey response level, showing 95% confidence intervals on this prediction from the estimated model. The chart is produced by running `margins, at(risk_attitude=(0(1)10)) predict(xb)`.

`risk_attitude`	**Margin**	**SE**	z	$P > \|z\|$	**[95% Conf. Interval]**
0	0.6037262	0.0439114	13.75	0.000	[0.5176614, 0.689791]
1	0.5659963	0.0369109	15.33	0.000	[0.4936523, 0.6383403]
2	0.5282664	0.0304803	17.33	0.000	[0.468526, 0.5880067]
3	0.4905364	0.0250625	19.57	0.000	[0.4414148, 0.5396581]
4	0.4528065	0.0214394	21.12	0.000	[0.4107861, 0.494827]
5	0.4150766	0.0205817	20.17	0.000	[0.3747372, 0.455416]
6	0.3773467	0.0228036	16.55	0.000	[0.3326524, 0.4220409]
7	0.3396167	0.027365	12.41	0.000	[0.2859824, 0.3932511]
8	0.3018868	0.0333185	9.06	0.000	[0.2365838, 0.3671898]
9	0.2641569	0.0400479	6.60	0.000	[0.1856644, 0.3426494]
10	0.226427	0.0472228	4.79	0.000	[0.1338719, 0.318982]

Table 4: Delta-method Margins

As the risk attitude score increases from 0 to 10, the predicted mean CRRA decreases steadily from approximately 0.60 to 0.23. Each unit increase in the risk attitude is associated with a decrease in the predicted CRRA, with 95% confidence intervals that are narrow and do not substantially overlap between successive scores. This pattern again indicates that higher self-reported willingness to take risks corresponds to lower estimated levels of risk aversion in the experimental setting, as well as supports the validation of the survey instrument by confirming that it effectively captures individuals' true risk preferences. The consistent inverse relationship between the survey risk attitudes and experimentally measured CRRA suggests that the instrument reliably predicts actual risk-taking behavior. Individuals who rate themselves as more risk-tolerant not only have lower CRRA estimates but also exhibit behavior in line with their self-assessments.

5 Extending the Post-Estimation Analysis

This section provides results from Task 4 of the assignment.

In preparation for the simulation process, I stored the estimated coefficients from the previous regression in scalars, and built an prediction model:

$$\texttt{pred\_mean} = \beta_0 + \beta_1 * \texttt{risk\_attitude}$$

I then simulated CRRA values for each `risk_attitude` level by generating 1000 random Normal draws for those values of the mean and variance, calculated the mean and 95% confidence intervals, and generated a table and a graph based on the simulation results, both of which are presented below:

Because the data generation process is random, the simulation is different each time you run the code file. I've only included an example, but please feel free to try my code.

risk_a~e	mean_c~a	p2_5	p97_5
0	.5940551	-.2334366	1.460884
1	.5577046	-.3121462	1.387828
2	.5290354	-.324188	1.42221
3	.4867336	-.3644379	1.307482
4	.4552341	-.3514957	1.301346
5	.4160047	-.3933557	1.201797
6	.4000891	-.4246299	1.260857
7	.3249075	-.5175887	1.138151
8	.324099	-.5437163	1.197112
9	.2834082	-.552037	1.179043
10	.2393259	-.6460882	1.077323

Figure 1: Prediction result from simulation, with 95% confidence interval

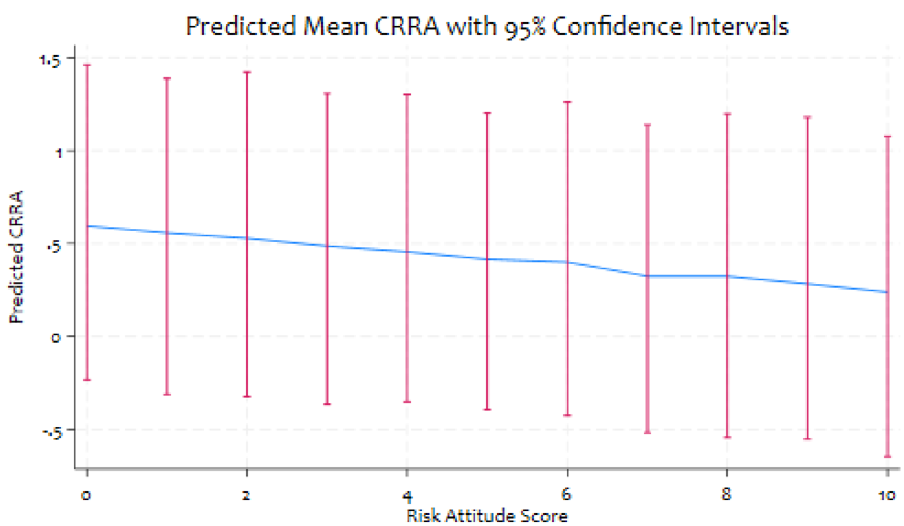

Figure 2: Prediction result from simulation, with 95% confidence interval, presented graphically

Based on the results from the data generation and simulation process, the conclusion about the validation of the survey instrument remains consistent and is further strengthened. As individuals report higher willingness to take risks (higher survey scores), the predicted mean CRRA decreases, indicating lower levels of risk aversion in the experimental measure. The 95% confidence intervals around these predicted means are narrow and generally do not overlap between successive risk attitude levels, suggesting that the differences are statistically meaningful.

6 Reflection

This section serves as the answer to Task 5 of the assignment.

The method in Task 3 provides evidence of this relationship through point estimates and tests of significance, suitable for inferences about population parameters. Specifically, this method is effective for estimating the average effect of the self-reported risk attitude on the experimentally measured risk aversion (CRRA). This method provides **point estimates** of the relationship, allowing you to quantify how changes in risk attitude scores are associated with changes in the mean CRRA values. It works well for hypothesis testing and making inferences about population parameters, such as determining whether the relationship between risk attitude and CRRA is statistically significant. It's better at delivering precise estimates of the regression coefficients and to assess the significance of these estimates using confidence intervals and p-values. However, it is **not as effective at capturing the variability or distribution** of CRRA values at each level of risk attitude. It assumes homoskedasticity and focuses on the central tendency, providing limited insight into how individual CRRA values may vary around the predicted mean.

The method in Task 4, certainly a more explorative and frequentist one, enriches this analysis by exploring the predicted distributions of CRRA across different risk attitude levels, highlighting how both the mean and variability of CRRA change with self-reported risk tolerance. It builds upon the results of the first by simulating the distribution of CRRA values for each risk attitude score. It is better at estimating the entire distribution of CRRA associated with each level of risk attitude. Specifically, it excels in illustrating how both the mean and the variability of CRRA change across different risk attitudes, providing 95% confidence intervals that reflect the uncertainty and dispersion of the predicted values. However, it is less focused on **estimating precise parameter values** and is not primarily designed for formal hypothesis testing about population parameters. That's because reliance on simulation introduces approximation, and depends on the assumption that the error terms are normally distributed, which may not hold in all cases. The simulation also introduces data that are not empirically procured, which, depends on the experiment, may yield inaccurate information.