Glenn Harrison: Homework Assignment

All of these questions involve the claim to have developed a short hypothetical survey instrument to measure risk aversion that is validated by comparison to a simple, incentivized experiment undertaken with the same subjects. This study is Dohmen, Thomas; Falk, Armin; Huffman, David; Sunde, Uwe; Schupp, Jürgen, and Wagner, Gert G., "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences," *Journal of the European Economic Association*, June 2011, 9(3), 522-550. Many researchers just take their word for it.

The raw data are in two files, dohmen.dta in *Stata* format, and the same data in dohmen.xlsx for anyone wanting to use the data in *R* or other statistical packages. The variable descriptions are listed at the end of the assignment. You can undertake all of the calculations in any statistical package you like. In all cases, just provide the command file that does everything (e.g., do not assume some data set is already read in), a text LOG file of all commands and their output, and a PDF document to report your results. To illustrate, an outline of the command file in *Stata*, imaginatively named Homework.do, is provided.

1. **Replicate** the statistical model estimated in the paper in Table 2, page 533. Focus on the results on column (1) of part (b). This is an interval regression model of the row at which the subject switched in the experimental task, with the survey risk attitude as a covariate. Once you have estimated the model, display the ML estimates, where ML stands for "Maximum Likelihood". If you do not know what an interval regression is, then look it up. Interval regression with *Stata* is explained in the Stata Manual, and I attach a PDF of it. It is also explained at https://stats.oarc.ucla.edu/stata/dae/interval-regression/, and at [https://stats.oarc.ucla.edu/stata/output/interval-regression/.](https://stats.oarc.ucla.edu/stata/output/interval-regression/) Interval regression with *R* is explained at [https://stats.oarc.ucla.edu/r/dae/interval-regression/.](https://stats.oarc.ucla.edu/r/dae/interval-regression/) In most respects it is just like OLS: you are estimating a mean and a standard deviation of some latent construct, and you end up with a point estimate and a standard error on the mean and a point estimate and a standard error on the standard deviation. The only difference is that you recognize that the data is defined by an interval. If you collapse the intervals to their mid-point and just run OLS, you fail the assignment, even if some journals allow such nonsense. When the lower and upper bounds for all data are the same, interval regression is just OLS. You may have clopen intervals, where one of the bounds is $+\infty$ or $-\infty$; this matters for the rest of the homework. Explain in words what the regression replication is saying in terms of the claim of validation. Note that many interval regression models report the estimates of the log of the variance, and it will be useful for you to refer to the estimates of the standard deviation. The R package referenced above does not report that estimated standard deviation.
2. Repeat the estimation but now **allow the survey risk attitude to affect the standard deviation** of the estimated CRRA. In the *Stata* command for interval regression this involves use of the heteroskedastic option. I have no clue how to do it in R, but this is

so standard it must in some package. Note that this is not asking you to test for heteroskedasticity, just to estimate a full model that allows for it. As a hint for question #4, and in general, this option is not just a statistical afterthought. Test for the *joint* significance of the covariates for the survey risk attitude.

3. **Repeat the estimation but now use the CRRA intervals** as the dependent variable, so we see if the survey response is related to a measure of risk preferences under EUT. Then show numerically and in a display how the predicted mean of CRRA varies with the hypothetical survey response level, showing 95% confidence intervals on this prediction from the estimated model. Draw carefully stated conclusions about the claim of validation of the survey instrument. Do not use words to just repeat what you have estimated: interpret it.
4. Now **extend the post-estimation analysis** of the interval regression model to do more than just show the association of the predicted mean of CRRA with the hypothetical survey response level. Repeat the exercise in step 3, but looking at the predicted CRRA, and 95% confidence intervals for it, and say if your conclusions about the claim of validation of the survey instrument remain the same. HINT: the interval regression models the latent variable CRRA, which has a mean and a variance. In step 3 you looked at the effect of the hypothetical survey response on the mean. Use the estimates from step 3 to predict the mean and variance for survey response 0. Then simulate 1000 random Normal draws for those values of the mean and variance (this is an approximation, but will do). Collect these simulated values, or just save the mean and 95% confidence interval for the simulated CRRA distribution. Then repeat for each of survey responses 1, 2, ..., 10. Use these simulated distributions to show in a display how the predicted mean of CRRA varies with the hypothetical survey response level, showing 95% confidence intervals on this prediction from the estimated model of the CRRA distribution.
5. **Say something in words** about what types of statistical inferences the results in steps 3 and 4 might be appropriate or not appropriate for. Again, forget what others do, and forget what you can sneak past referees and editors at better academic journals.

As a general matter, risk preferences in the experimental literature are rarely modeled using interval regression. It has some value in exercises with data from Multiple Price Lists, and requires that one estimate risk preferences with just one parameter, which usually restricts the models of risk preferences to either Expected Utility Theory or Dual Theory. In general one needs to estimate customized maximum likelihood or Bayesian models of structural theories of risk preferences, as explained by Harrison, Glenn W., and Rutström, E. Elisabet, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison (eds.), *Risk Aversion in Experiments* (Bingley, UK: Emerald, Research in Experimental Economics, Volume 12, 2008) and Gao, Xiaoxue Sherry; Harrison, Glenn W., and Tchernis, Rusty, "Behavioral Welfare Economics and Risk Preferences: A Bayesian Approach," *Experimental Economics*, 26, 2023, 273-303, respectively. Harrison and Rutström (2008) explain the historical use of Multiple Price Lists and interval regression in detail.

```
Contains data from dohmen.dta
 Observations:           450
    Variables:            26                  29 Oct 2024 06:45
--------------------------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
----------------------------------------------------------------------------------------------
r1safe          byte    %12.0g     mpl        Switching Row to Safe Payment
female          float   %9.0g                 1=female
age             float   %9.0g                 Age in Years
height          float   %9.0g                 Height in cm
middleschool    float   %9.0g                 Middle School (Realschule)
abitur          float   %9.0g                 High-School Degree (Abitur)
no_degree       float   %9.0g                 No Schooling Degree
other_degree    float   %9.0g                 Other Schooling Degree
enrolled_inschool
                float   %9.0g                 Currently Enrolled in School
child1          float   %9.0g                 1 Child
child2          float   %9.0g                 2 Children
child3          float   %9.0g                 3 or more Children
loghhnet        float   %9.0g                 Log Household Monthly Income (Net of Taxes and Benefits)
risk_attitude   byte    %8.0g      f019       Willingness to Take Risks (0-10)
row_hi          float   %9.0g                 Upper row number of the switch
row_lo          float   %9.0g                 Lower row number of the switch
record          byte    %8.0g                 Data record number
countI          float   %9.0g                 Number of subjects in Metric I
row_mid         float   %9.0g                 Midpoint of the row where the switch occurred
ce_lo           float   %9.0g                 Lower certainty equivalent at the switch
ce_hi           float   %9.0g                 Upper certainty equivalent at the switch
ce_mid          float   %9.0g                 Midpoint of the CE where the switch occurred
countII         float   %9.0g                 Number of subjects using Metric II
crra_lo         float   %9.0g                 Lower bound of CRRA
crra_hi         float   %9.0g                 Upper bound of CRRA
crra_mid        float   %9.0g                 Mid-point of CRRA intervals
```