

University of Central Florida
College of Business

QMB 6912
Capstone Project in Business Analytics

Problem Set #10

Due Date: Sunday, 16 April 2023, at 11:59 PM.

In this problem set, you are to estimate a sample-selection model for the prices of houses sold as either owner-occupied homes or rental properties. Following the examples discussed in class, begin with a probit model to predict the probability that a house will be sold to a particular type of buyer. Use these variables as the starting point for the selection equation in your empirical specification of your sample selection model.

Next, collect the variables that you found relevant for predicting the prices of homes sold to the two types of buyers. Use these as the starting point for the specification of your observation equations. Trim down your model until all the coefficients are significant, just as you have done in previous problem sets.

During your estimation, you might notice that the estimation algorithm does not converge or that some parameters are unable to be estimated. First, adjust some of the hyper-parameters, such as the number of iteration in the optimization routine. Next, try changing your specification by removing some of the variables from your model. As shown in the examples in the R documentation, it often helps to reserve some variables for the selection equation to differentiate from the observation equation. If all your attempts fail, use the results from Heckman's two-step estimator as an approximation.

Finally, present the results of your estimates of the sample selection model in L^AT_EX tables. Then, write a brief summary of your interpretation of the empirical results. Be sure to compare the results with those from the modeling approaches in the previous problem sets and note any differences in the values of houses when selection into the types of home buyers is taken into account.

Prepare and compile your work in \LaTeX and include scripts for any of the calculations in R. In particular, create the following directory structure, separate from your existing work:

- Code/
- Data/
- Figures/
- Tables/
- Text/
- Paper/
- Misc/

In a file called `README.md`, which should also live in the directory containing the above folders, provide the instructions concerning how to run the executable shell script `DoWork.sh` (in the same directory) that will execute the code that produced all of the answers collected and documented in your report, which will live in the subdirectory `Paper/`. In the subdirectory `Code/`, keep the R code; in `Data/` keep the raw data file you downloaded, so that `DoWork.sh` can load it into R, and in `Figures/` keep any figures you created for your answers. Similarly, keep any \LaTeX scripts for tables in the `Tables/` folder. You may put any written text in the `Text/` folder, if not already included in a `tex` file in your `Paper/` folder. Put anything else in the subdirectory `Misc/`. I should then be able to replicate all of your work simply by typing

- `$./DoWork.sh`

on the command line of a terminal window.

To provide you a template, which makes preparation easier for you and grading easier for me, I have placed sample \LaTeX and R code in the GitHub repository for the course: `QMB6912S23`, under my GitHub username `LeeMorinUCF`; pull this repository and use these files a framework within which to create the answers for this problem set. Push the files to a folder on your GitHub repository and I will pull your submissions to my computer for grading.

Be sure to support your calculations with descriptions of what you were trying to do (for example, in comments in your R code as well as in the \LaTeX explanations) because partial credit will be given.

Due Date: Sunday, 16 April 2023, at 11:59 PM.