Beginning Notes

1. Deadline: 1 week after receiving the task. Estimated time for completion is 12-15 hours.

2. Selection is merit-based and on a rolling basis.

3. Please write up your results in the format of a report in English.

Data Description

Conceptual Understanding:

Our data is from an online real estate marketplace, which is similar to "贝壳网" in China or Homesnap.com in the US. Generally, sellers post their properties on the online platform, and buyers would make offers and bargain with the sellers. We, as researchers, are interested in studying the bargaining process.

Definition of a Bargaining Event:

The typical process of a bargaining event is as follows. A seller lists her house on the platform, so her listing can be seen by potential buyers who are looking to buy a property. The buyers can choose to send their offers to the seller through the online platform. The seller can either accept the offer or revise the listing price. Rounds of bargaining continue until the property is taken off the market by the seller, which could happen under two scenarios: 1) The seller accepts only one offer and the property is automatically taken off the market; 2) The seller takes the property off the market. He or she rejects some buyers and goes through a private negotiation (not available in our data) with the remaining buyers. The deal succeeds if the seller accepts one of the offers from the remaining buyers. The deal fails if the seller accepts no offers.

The following figure illustrates a typical bargaining event:
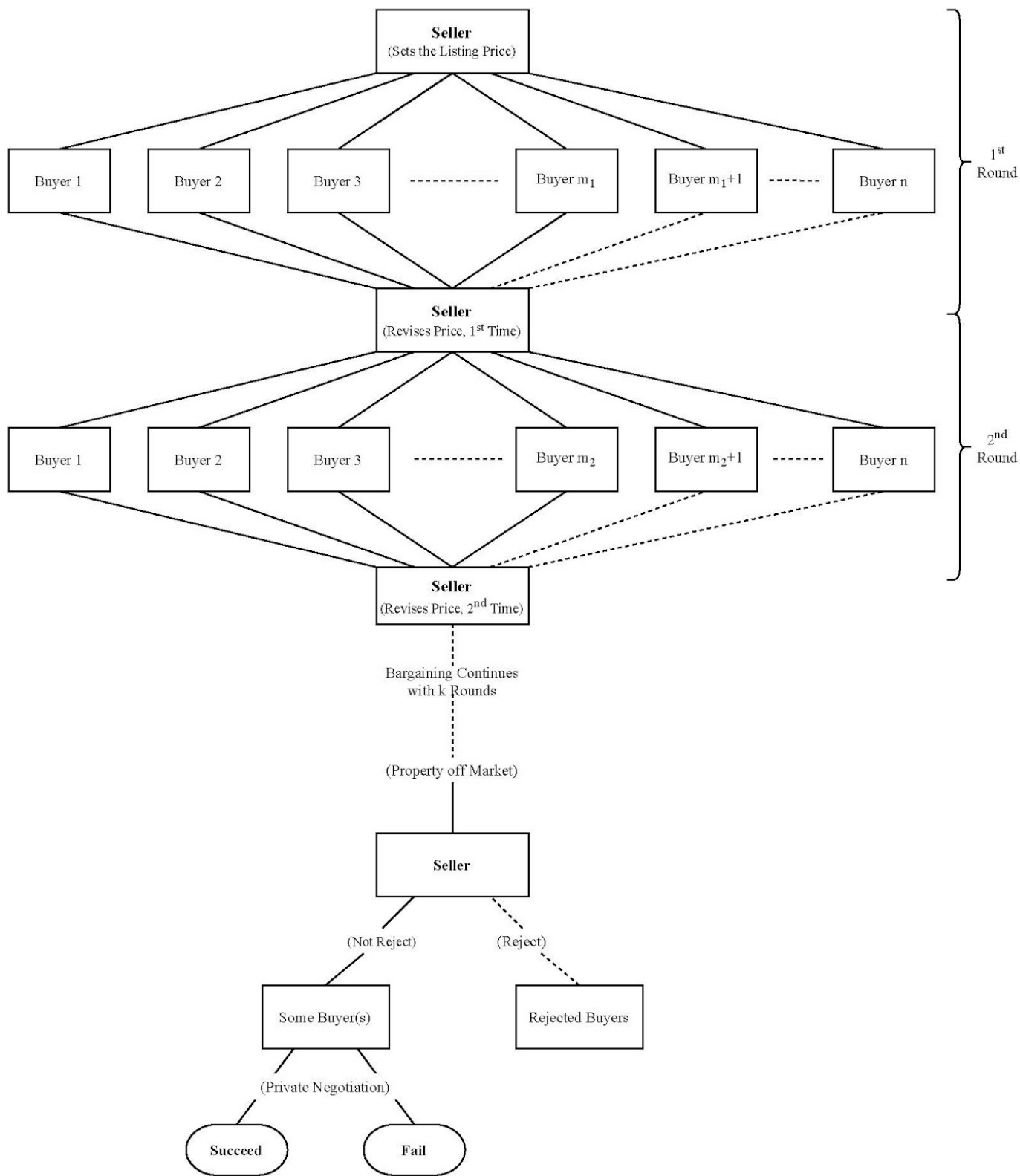
Data Structure:

The provided data set records the above bargaining process of sales of homes in the U.S. housing market. Each row is an action by a seller or a buyer during a bargaining event. For example, a seller listing her property can be an action; a buyer making an offer can also be an action. The person who takes the action is recorded by the "actor" column (could be either "buyer" or "seller"). Each bargaining event is identified by "event_id", and each buyer is identified by "buyer_id".

In the package, you will find the following files:

1. "selected_bargaining_data.csv" is the main dataset you will be working with.

2. "Data Structure.xlsx" illustrates three example events in your data.

3. "codebook.xlsx" includes the codebook to all the key variables.

4. "cbsa2fipsxw.csv" contains the FIPS code for each state. It might be helpful when you are working on task 2 (drawing a map).

Tasks

You can use any statistical software of your choice such as Stata, Python, or R.

1. Produce a summary statistics table at **event level** (note that our current data is at action-level). The table should include mean, standard deviation, minimum, median, and maximum. The variables to be summarized include property age (using the variable "built_year"), number of buyers (identified by unique "buyer_id"), number of seller revisions, duration until off market (using variable "off_market_date", in days), and sales price (using the variable "de_sales_price").

    Note: Remember to identify and exclude anomalies or extreme values when you calculate summary statistics for these variables; use the deflated price rather than the raw price.

    Here is an example of the desired format of the summary statistics table, and please write code that outputs the results into a single file (e.g., csv, txt, etc.).

	Mean	Std.Dev.	Min	Median	Max
Property Age	35.17	32.01	1	41	179
Buyer Represented by the Platform	1.23	0.25	1	2	4
Revisions	0.85	1.27	0	3	12
Duration Until Off-Market (Days)	31	25	1	19	82
Sales Price	459,307	252,942	121,367	402,598	1,498,686

2. Draw a map with all states in the U.S. to illustrate the geographical distribution of the bargaining events in the data sample. Use colors to represent the number of bargaining events from each state in our sample. A clear legend should be provided.

    Hint: In order to locate each event, make use of the "censustract" variable in the dataset. This variable contains the "FIPS state code" which uniquely identifies which state the event was located. If it has 11 digits in total, the first two digits represent the state's FIPS code; if it has 10 digits, then the first digit will be the state's FIPS code. Once you get the state's FIPS code, make use of the "cbsa2fipsxw.csv" file which connects the state's FIPS code to the corresponding state's name.

3. What factors can affect the final sales price of a property ("sales_price" in the data)?

   Hints. You can run univariate or multivariate linear regressions to examine which variables are correlated with the final sales price of a property. Are these correlations statistically significant and economically large? Some potential factors to consider include *but are not limited to* initial listing price and number of buyers ("total_num_buyers_event" in the data). Maybe also consider taking some transformation of the price variables for the purpose of running regressions.

   Note. This analysis should be done at the event-level; use the deflated price rather than the raw price (variable begin with "de").

4. Interpret and compare the regression coefficients, and summarize your takeaway from 3.

5. Summarize all results in a report format.

6. **Bonus Question (Optional):**

   What are the possible confounding variables for the regressions you run in 3, and how could you mitigate the omitted variable bias problem? A conceptual discussion is enough.

   Hints. Suppose you want to test whether houses with a higher initial listing price are sold for higher prices, is it important that the houses you compare are similar in other aspects? If yes, how could you make the houses you compare more similar in these other aspects?

Submission

Please submit your code, output files (e.g., map, output file of summary statistics, log or output files of regressions) and the final report (in PDF) in a single compressed file to Franklin Qian at qzjquantum@gmail.com. Candidates will be considered on a rolling basis.