

# Econ 761 – Fall 2020

## Homework 5

Arjun Jain

Note: Tables I, II, and III are presented after explanations and answers to questions because the tables are large. Finally, code is attached at the end of the document.

### 1 Probit regressions of entry

For the Probit regression of Walmart entry on market level characteristics, I include the following variables: Kmart presence, log of county population, log of retail sales, percent of urban population, log of distance to Benton, Southern dummy, number of small stores, distance weighted number of Kmart stores.

For the Probit regression of Kmart entry on market level characteristics, I include the following variables: Walmart presence, log of county population, log of retail sales, percent of urban population, Midwest dummy, log of distance to Benton, Southern dummy.

This model does not take into account entry of a Walmart (or Kmart) into a county that already has a Walmart (or Kmart). In other words, one of the underlying assumptions is that each county has a maximum of one Walmart and one Kmart. It is reasonable to think pre-existing Walmarts may affect entry of an additional Walmart in any given county.

Another limitation of this model is that we are not exactly measuring entry of Walmart or Kmart. Rather, we have the data on presence of both large stores and are using this as a proxy for entry, which may not be the best representation.

### 2 Probit regressions with instrumenting strategy

Now, I use the variables excluded in question 1 as instruments in this question (for entry of Walmart and entry of Kmart), including the three optional unnamed variables. Essentially, we estimate a two-stage Probit model using excluded variables as instruments for entry.

The variable “distance from Benton county” was used as a covariate in each regression from question 1. Furthermore, it doesn’t seem very appropriate to use as an instrument for Walmart entry. This would mean we think “distance from Benton county” only affects Kmart entry through presence of Walmart. However, “distance from Benton county” probably has no effect on Walmart entry for a majority of counties considered. It would be interesting to know how Benton county in particular was chosen instead of another county, as the choice may just be random.

### 3 Bresnahan and Reiss analysis of industry

For this question, we estimate two ordered Probit models: the first using number of large players as the dependent variable, the second using total number of small and large players as the dependent variable.

A limitation here is again that we don't actually know the number of large players. We only know whether or not a particular county has a Walmart or Kmart. This affects our ability to accurately pin down both dependent variables used - number of large players and total number of small and large players.

### 4 Two-step method proposed by Bajari et al. (2012)

Now we use a two-step method. Using entry of Walmart as the dependent variable in the first stage Probit regression, we use the covariates determined to be part of the best specification in the first question. Then we can calculate the predicted value for Walmart entry, which we denote as  $\hat{Walmart}$ . This predicted value is then used in the second stage Probit regression, using entry of Kmart as the dependent variable and including covariates determined in question 1.

We can do the same process but using Kmart entry in the first stage. Then we can calculate the predicted value for Kmart entry, which we denote as  $\hat{Kmart}$ . This predicted value is then used in the second stage Probit regression, using entry of Walmart as the dependent variable and including covariates determined in question 1.

What we see from this simultaneous move game between Walmart and Kmart is that the presence of the other big player affects entry to a much larger degree than in the previous parts. Namely, there is a significant large negative effect of presence of Kmart on entry of Walmart, and vice versa.

## Tables

The following table is for question 1.

Table I		
	Walmart entry	Kmart entry
constant	-13.445 (1.110)	-2.738 (0.089)
Kmart	-0.739 (0.114)	– –
Walmart	– –	-0.556 (0.114)
county	– –	– –
log(county_pop)	2.282 (0.117)	1.752 (0.112)
log(retail_sales)	1.600 (0.127)	1.683 (0.142)
percent_urban	0.983 (0.207)	1.316 (0.231)
Midwest	– –	1.184 (0.173)
log(dist_Benton)	-1.091 (0.079)	0.682 (0.100)
South	0.870 (0.091)	0.922 (0.172)
small_stores	-0.084 (0.020)	– –
dist_Kmart	-3.880 (0.639)	– –
dist_Walmart	– –	– –

Note: in Table I, all twelve of the variables that aren't optional are included. However, not all of them are used in the specification for Walmart entry or Kmart entry. For example, “county” isn't included in either Probit. These covariates which are not included in a specification are indicated as such with “–”.

The following table is for question 3.

Table II		
	Number of large players (dependent variable)	Total number of small and large players (dependent variable)
log(county_pop)	2.019 (0.086)	1.865 (0.079)
log(retail_sales)	1.611 (0.102)	1.277 (0.086)
percent_urban	1.090 (0.170)	1.064 (0.154)
log(dist_Benton)	-0.390 (0.051)	-0.345 (0.048)
South	0.455 (0.071)	0.401 (0.065)
small_stores	-0.061 (0.016)	2.787 (0.071)
dist_Kmart	-2.560 (0.511)	-2.430 (0.485)

Note: in Table II, I only included the variables that are used as covariates in question 1. These are the coefficient estimates and standard errors that result from the ordered Probit.

The following table is for question 4.

	Walmart first stage, Kmart second stage	Kmart first stage, Walmart second stage
constant	-42/538 (2.271)	-31.686 (1.995)
$\hat{K}mart$	— —	-5.756 0.393
$\hat{W}almart$	-5.798 0.419	— —
log(county_pop)	4.000 (0.221)	4.084 (0.205)
log(retail_sales)	3.508 (0.217)	3.269 (0.205)
percent_urban	3.021 (0.292)	2.225 (0.252)
Midwest	1.620 0.196	— —
log(dist_Benton)	-0.064 (0.118)	-1.163 (0.092)
South	2.096 (0.213)	0.938 (0.103)
small_stores	— —	-0.120 0.022
dist_Kmart	— —	-4.410 0.688

Note: in Table III, I only included the variables that are used in one (or both) of the second stage regressions. I determined which variables to use as covariates in each stage depending on which ones were used in the best specification in the first question. For example, “number of small stores” isn’t included in the second stage regression for entry of Kmart, and so it is indicated by “—”.

```

1 // code for questions 2 and 3 of hw2
2
3 // clear workspace
4 clear
5
6 // import data
7 infile cnty log_cnty_pop log_cnty_rtl perc_pop midwest log_dist_B south kmart walmart num_stores
  dist_kmart dist_walmart opt1 opt2 opt3 using "C:\z_toshiba\course work\phd\econ 761\hw\hw5\XMat.out"
8
9 //1. probit regressions of entry
10 // Walmart entry (all variables included)
11 probit walmart kmart cnty log_cnty_pop log_cnty_rtl num_stores midwest south perc_pop log_dist_B
  dist_kmart dist_walmart
12
13 // Walmart entry (specifications that best fit the data)
14 probit walmart kmart log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart
15
16 // Kmart entry (all variables included)
17 probit kmart walmart cnty log_cnty_pop log_cnty_rtl num_stores midwest south perc_pop log_dist_B
  dist_kmart dist_walmart
18
19 // Kmart entry (specifications that best fit the data)
20 probit kmart walmart log_cnty_pop log_cnty_rtl midwest south perc_pop log_dist_B
21
22 //2. probit regressions with instrumenting strategy
23 // Walmart entry
24 ivprobit walmart log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart (kmart =
  cnty midwest dist_walmart opt1 opt2 opt3)
25
26 // Kmart entry
27 ivprobit kmart log_cnty_pop log_cnty_rtl midwest south perc_pop log_dist_B (walmart = cnty num_stores
  dist_kmart dist_walmart opt1 opt2 opt3)
28
29 //3. Bresnahan and Reiss analysis of industry
30 gen large_stores = kmart+walmart
31 gen total_stores = num_stores+large_stores
32
33 // dependent variable = number of large players
34 oprobit large_stores log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart
35
36 // dependent variable = total number of small and large players
37 oprobit total_stores log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart
38
39 //4. two-step method from Bajari et al. (2012)
40 // Walmart first stage, Kmart second stage
41 // first stage: find estimates using probit regression from question 1
42 probit walmart kmart log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart
43 predict walmart_hat
44
45 // second stage: use predicted value of Walmart entry in regression of Kmart entry
46 probit kmart walmart_hat log_cnty_pop log_cnty_rtl midwest south perc_pop log_dist_B
47
48 // Kmart first stage, Walmart second stage
49 // first stage: find estimates using probit regression from question 1
50 probit kmart walmart log_cnty_pop log_cnty_rtl midwest south perc_pop log_dist_B
51 predict kmart_hat
52
53 // second stage: use predicted value of Walmart entry in regression of Kmart entry
54 probit walmart kmart_hat log_cnty_pop log_cnty_rtl num_stores south perc_pop log_dist_B dist_kmart
55

```