

**Problem 1.**

*Solution:*

We use the following variables: log of the population, log of retail sales per capita, the urban ratio, a dummy for whether the county is in the Midwest, a dummy for whether the county is in the South, and the entry decision by the other firm ( $\delta_i$ ). We see that both firms have positive coefficients for all county-level variables. Furthermore, the presence of a rival firm reduces firm  $i$ 's probability of entry. The main limitation of this model is the fact that the entry decision of each firm is endogenous and should not be thought of as an exogenous regressor. Walmart's decision to enter depends on Kmart's entry decision, but simultaneously Kmart's entry decision depends on Walmart's entry decision. As the entry decisions of either firm are simultaneously determined, we cannot separately regress Walmart's probability of entry on Kmart's entry decision and vice versa, since each should affect the other if we assume firms are behaving strategically.  $\square$

Table I: Probit Regressions, Walmart and Kmart

	<i>Dependent variable:</i>	
	Walmart	Kmart
	(1)	(2)
Log of population	1.647*** (0.081)	1.747*** (0.110)
Log of retail sales per capita	1.344*** (0.115)	1.723*** (0.147)
Urban ratio	1.235*** (0.185)	1.238*** (0.229)
Midwest	0.958*** (0.135)	0.578*** (0.148)
South	1.543*** (0.137)	0.328** (0.148)
$\delta_i$	-0.817*** (0.109)	-0.710*** (0.113)
Constant	-17.554*** (1.018)	-21.536*** (1.377)
Observations	2,065	2,065
Log Likelihood	-788.029	-566.926
Akaike Inf. Crit.	1,590.059	1,147.853

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Problem 2.**

*Solution:* We now consider whether an instrumenting/control function approach may be useful. Suppose we do a probit regression of Walmart's probability of entry on the distance

from Benton County (home of Walmart headquarters) and then use the predicted probability of Walmart entering in our probit regression for Kmart's entry decision. This could be a valid approach, since the distance of a given county from Benton county arguably affects the profitability of opening a Walmart location in the county. If the county is very close to Benton County, the logistical cost of opening a new location in that county is likely to be low (due to low cost of transportation, ease of corporate travel to and from the location, etc.). Conversely, if the new county is very far from Walmart headquarters, Walmart might face significantly more logistical hurdles if it were to build a new location there (increased time/cost of transportation, less seamless corporate integration, etc.). In summary, we can plausibly argue that the distance of the given county from the Walmart headquarters in Benton County affects the profitability of entry and thus impacts Walmart's entry decision. In order for this to be a valid instrument, however, the distance from Benton county must also not impact the profitability of Kmart entry. This is reasonable, since Kmart at the time was headquartered in Troy, Michigan, meaning that the distance of a county from Benton county likely has little (if any) first order impact on Kmart's logistics. It only indirectly affects Kmart's entry insofar as it affects Walmart's entry, but this is permissible.  $\square$

### Problem 3.

*Solution:* We now estimate two "Bresnahan and Reiss" ordered Probit models. The first specification is an ordered Probit regression of the number of large players on market-level characteristics, whereas the second specification is an ordered Probit on the total number of small and large players in the market (where small players are small retail stores). The estimates for each specification are included in Table II: One thing we note here is that the

Table II: Ordered Probit Regressions

	<i>Dependent variable:</i>	
	of large stores	Total of stores
	(1)	(2)
Log of population	1.775*** (0.070)	1.388*** (0.046)
Log retail	1.527*** (0.098)	0.828*** (0.060)
Urban ratio	1.310*** (0.162)	-0.746*** (0.114)
Midwest	0.855*** (0.120)	-0.193** (0.087)
South	1.125*** (0.118)	0.656*** (0.086)
Observations	2,065	2,065

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

coefficients for the regression with large stores only are broadly consistent with the results

in Table II, albeit with some small differences. The coefficients are more different when we consider the ordered Probit model with both small and large players. Here, we see that the coefficient on the urban ratio and midwest are negative. Furthermore, even the positive coefficients are smaller than the ones in previous regressions. What is going on here? The answer lies in the homogeneity assumption of the ordered Probit regression. The ordered Probit regression model assumes that the identity of the firms is irrelevant for the entry decision (since we are regressing only on the total number of firms). This means a county with two small, local retail stores has the same realized level of the dependent variable as a county with both a Walmart and a Kmart. However, counties which only have two small local retailers are likely to be very different in characteristics from a county with two major retail operations! The reason why the urban ratio coefficient becomes negative is likely the following: small counties with a much smaller urban population may be more likely to have a bunch of small, local retail stores. This correlates to a relatively high number of total firms. Conversely, counties with a much larger urban population may have a Walmart and only a few local retail stores (because Walmart is only likely to enter if there is a significant urban population, and the entry of Walmart may have led smaller retail stores to exit). In the presence of this relationship, our regression model would find that the total number of retail stores is negatively associated with the urban ratio.

A problem with both models is the homogeneity assumption. Walmart is different from Kmart, and both are very different from small retail stores. For a given market, the entry decision of Walmart will be very different from the entry decision of a small retailer (holding competitors constant) due to differences in cost structure, logistical networks, scale, productivity, etc. Furthermore, Walmart's entry decision is going to be dramatically different if their opponent is Kmart versus a small retail store. Similarly, the entry (and exit) decisions of small retail stores are going to differ based on whether they are competing with two other small retail stores versus competing with a Walmart and a Kmart. In summary, there are many limitations in using an ordered Probit model to model the strategic considerations of firms.  $\square$

#### Problem 4.

*Solution:* We now aim to add heterogeneity to the model. Suppose that Walmart and Kmart play an entry game of incomplete information. The profit functions to firm  $i$  in market  $m$  given action  $y$  is given by the following:

$$\pi_{im}(y) = y_{im}(x_{im}\beta_i + \Delta_i y_{-i} + \varepsilon_i)$$

where  $\beta_i$  and  $\Delta_i$  are parameters,  $x_{im}$  are firm-market level observables,  $y_{-i}$  is the entry decision of the other firm, and  $\varepsilon_i$  is the unobservable part of profits. Suppose that  $\varepsilon_i \sim N(0, 1)$ . Let  $y_{im} = 1$  indicate that firm  $i$  enters market  $m$ , and  $y_{im} = 0$  indicate that firm  $i$  chooses not to enter  $m$ . This implies the following:

$$\pi_{im}(1) = x_{im}\beta_i + \Delta_i y_{-i} + \varepsilon_i$$

$$\pi_{im}(0) = 0$$

We will use the framework for Bajari et al. to estimate  $\beta_i$  and  $\Delta_i$ . We will have  $x_{im}$  be the same county-level covariates used in previous regressions, with the inclusion of the log

of distance from Benton County for Walmart. Let  $\sigma_i(y_i = k \mid x_{im})$  indicate the probability that player  $i$  chooses  $y_i = k$  given state variables  $x_{im}$ . The expected payoff  $\Pi_i$  to player  $i$  is obtained by the following:

$$U_i(y_{im}, x_{im}, \varepsilon_i; \beta) = \sum_{y_{-i}} \pi_{im}(y_{im}, y_{-i}, x_{im}, \beta) \sigma_{-i}(y_{-im} \mid x_{im}) + \varepsilon_i$$

We can also define the deterministic component of  $i$ 's payoffs as the following:

$$\Pi_i(y_{im}, x_{im}, \varepsilon_i; \beta) = \sum_{y_{-i}} \pi_{im}(y_{im}, y_{-i}, x_{im}, \beta) \sigma_{-i}(y_{-im} \mid x_{im})$$

We can also deduce that the optimal action for player  $i$  must satisfy the following:

$$\sigma_i(y_{im} \mid x_{im}) = P(U_i(y_{im}, x_{im}, \varepsilon_i; \beta) > U_i(y_{jm}, x_{im}, \varepsilon_i; \beta))$$

Let  $\sigma_w(1 \mid x_m)$  indicate the probability Walmart enters market  $m$  given market-level state variables  $x_m$ . It follows that  $\sigma_w(0 \mid x_m) = 1 - \sigma_w(1 \mid x_m)$  represents the probability Walmart doesn't enter. We define  $\sigma_k(1 \mid x_m)$  and  $\sigma_k(0 \mid x_m)$  similarly. We do not know these probabilities, but we may have estimates  $\hat{\sigma}_w(1 \mid x_m)$  and  $\hat{\sigma}_k(1 \mid x_m)$ . These estimates may be obtained parametrically or non-parametrically. We will estimate them parametrically using Probit regressions. Specifically, we do a probit regression of Walmart's entry based on the county level covariates in part 1 as well as the log of the distance from Benton County (we do not regress on Kmart's entry). Similarly, we estimate a Probit regression Kmart's entry decision on the county level covariates in part 1 (not including Walmart).

Given these estimated probabilities of entering, we can determine the payoffs to each firm of entering vs staying out. The expected utility of firm  $i$  (with opponent  $j$ ) from entering market  $m$  is given by

$$\begin{aligned} \hat{\Pi}_i(1, x_{im}) &= \hat{\sigma}_j(1, x_{jm})(x_{im}\beta_i + \Delta_i + \varepsilon_i) + (1 - \hat{\sigma}_j(1, x_{jm}))(x_{im}\beta_i + \varepsilon_i) \\ &= x_{im}\beta_i + \hat{\sigma}_j(1, x_{jm})\Delta_i + \varepsilon_i \end{aligned}$$

We also find that

$$\hat{\Pi}_i(0, x_{im}) = 0$$

Given that  $\varepsilon_i \sim N(0, 1)$ , we predict that Player  $i$  will enter if and only if

$$\begin{aligned} \Delta \hat{\Pi}_i(1, 0, x_{im}, \beta) &\geq 0 \\ \iff \hat{\Pi}_i(1, x_{im}, \beta) &\geq 0 \\ \iff x_{im}\beta_i + \hat{\sigma}_j(1, x_{jm})\Delta_i + \varepsilon_i &\geq 0 \\ \iff \varepsilon_i \geq -x_{im}\beta_i - \hat{\sigma}_j(1, x_{jm})\Delta_i \\ \iff \varepsilon_i \leq x_{im}\beta_i + \hat{\sigma}_j(1, x_{jm})\Delta_i \end{aligned}$$

where the last line follows from the symmetry of the Normal distribution. It then follows that

$$\begin{aligned} \hat{\sigma}_i(1, x_{im}) &= P(\Delta \hat{\Pi}_i(1, 0, x_{im}, \beta) \geq 0) \\ &= P(\varepsilon_i \leq x_{im}\beta_i + \hat{\sigma}_j(1, x_{jm})\Delta_i) \\ &= \Phi(x_{im}\beta_i + \hat{\sigma}_j(1, x_{jm})\Delta_i) \\ &= \Phi(\hat{\Pi}_i(1, x_{im}, \beta)) \end{aligned}$$

Hence,

$$\hat{\Pi}_i(1, x_{im}) = \Phi^{-1}(\hat{\sigma}_i(1, x_{im})) = x_{im}\beta_i + \Delta_i\hat{\sigma}_j(1, s) + \varepsilon_i$$

In other words, we can back out the expected entry profits implied by the estimated entry probability using the inverse quantile function of the Normal distribution. Thus, in order to estimate  $\beta_i$  and  $\Delta_i$ , we can simply regress  $\Phi^{-1}(\hat{\sigma}_i(1, x_{im}))$  on  $x_{im}$  and  $\hat{\sigma}_j(1, s)$  using Weighted Least Squares. We include the results of our regression below:

Table III: Two-stage semi-parametric WLS, Probit shocks

	<i>Dependent variable:</i>	
	Walmart (1)	Kmart (2)
Log population	1.823*** (0.003)	1.758*** (0.011)
Log retail	1.544*** (0.002)	1.730*** (0.006)
Log Benton distance	-1.079*** (0.002)	
Urban ratio	1.306*** (0.004)	1.236*** (0.020)
Midwest	0.002 (0.002)	0.580*** (0.018)
South	0.702*** (0.002)	0.340*** (0.018)
$\Delta_i$	-0.714*** (0.022)	-0.764 (0.023)
Constant	-12.236*** (0.014)	-21.612*** (0.052)
Observations	2,065	2,065
R <sup>2</sup>	0.997	0.995
Adjusted R <sup>2</sup>	0.997	0.995
Residual Std. Error	1.262 (df = 2057)	1.471 (df = 2058)
F Statistic	115,641.400*** (df = 7; 2057)	71,692.330*** (df = 6; 2058)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

We find that the direction and magnitude of the coefficients is similar to what we have found previously in Table I, albeit with some slight differences. Namely, the coefficient estimates for each county-level covariate seems to be higher for Walmart than the initial regression in Table I, whereas the impact of KMart entry appears to be smaller in absolute value. The coefficient estimates for Kmart do not differ significantly from those found in Table I.

Alternatively, based on the approach outlined in section 4.2 of the paper, we can form moments using the observed choices and the choice probabilities ( $\sigma(1, \hat{\sigma}_k, x_{wm})$ ) implied by the first stage estimated probabilities ( $\hat{\sigma}_k$ ) and parameter vector  $\theta_i$ . Specifically, we form the moment vectors

$$g_w(\theta) = [y_{wm} - \sigma(1, \hat{\sigma}_k, x_{wm}, \theta_w)]_{m=1}^M$$

$$g_k(\theta) = [y_{km} - \sigma(1, \hat{\sigma}_w, x_{wm}, \theta_k)]_{m=1}^M$$

where  $\theta_w = (\beta_w, \Delta_w)$  and  $\theta_k = (\beta_k, \Delta_k)$ . Assuming a probit specification, we have that

$$\sigma(1, \hat{\sigma}_k, x_{wm}, \theta_w) = \Phi(x_{wm}\beta_w + \Delta_w\hat{\sigma}_k(1 \mid x_m))$$

Given these moment vectors, we can construct sample moment vectors and use GMM. We use the WLS coefficient estimates as the initial value for our GMM estimation. We list our coefficients in Table IV alongside our WLS estimates. They are pretty similar to our WLS coefficient estimates, which is mildly reassuring. GMM seems to find smaller values for  $\Delta_i$ .

Table IV: Two-stage semi-parametric WLS and GMM, Probit shocks

	Walmart:		KMart:	
	WLS	GMM	WLS	GMM
	(1)	(2)	(3)	(4)
Log population	1.823*** (0.003)	1.887	1.758*** (0.011)	1.810
Log retail	1.544*** (0.002)	1.550	1.730*** (0.006)	1.666
Log Benton distance	-1.079*** (0.002)	-1.152		
Urban ratio	1.306*** (0.004)	1.540	1.236*** (0.020)	1.925
Midwest	0.002 (0.002)	0.061	0.580*** (0.018)	1.014
South	0.702*** (0.002)	0.682	0.340*** (0.018)	0.132
$\Delta_i$	-0.714*** (0.022)	-0.400	-0.764 (0.023)	-0.673
Constant	-12.236*** (0.014)	-12.166	-21.612*** (0.052)	-21.728

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

To compute standard errors, I would use the fact that, if  $N$  is the sample size,

$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow_d N(0, (\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma})^{-1})$$

where

$$\hat{\Gamma} = \frac{\partial E[g(\theta)]}{\partial \theta'} \Big|_{\theta=\hat{\theta}}$$

and

$$\Omega = E[g(\hat{\theta})g(\hat{\theta})']$$

We then have that  $NVar(\hat{\theta}) \rightarrow_p (\hat{\Gamma}'\hat{\Omega}^{-1}\hat{\Gamma})^{-1}$ . We can use the diagonal elements of the variance covariance matrix to construct the standard errors.

Finally, if we want to go full parametric, we can do a probit regression of entry based on market level covariates and our first-stage predicted probabilities of entry. I have included the results for this specification in Table V. Of course, the coefficients are different due to the fact that we are assuming a Probit regression model, and the interpretation of Probit coefficients is different than that of least squares (since the marginal impact of a regressor depends on the value of each of the other regressors, meaning that we don't have the same natural interpretation as we do with OLS where the coefficients are the marginal change in the dependent variable). Also, it is well known that the scale parameter for Probit regression is not identified, so the magnitude of these coefficients is not uniquely identified. Nonetheless, the sign of each of the coefficients agrees with what we have found previously.

□

Table V: Fully parametric two-stage Probit

	<i>Dependent variable:</i>	
	Walmart	Kmart
	(1)	(2)
Log population	4.802*** (0.230)	4.635*** (0.254)
Log retail	4.352*** (0.247)	3.828*** (0.233)
Log Benton dist	-1.462*** (0.120)	
Urban ratio	3.648*** (0.289)	3.211*** (0.301)
Midwest	0.579*** (0.181)	1.966*** (0.186)
South	1.039*** (0.182)	2.705*** (0.230)
$\Delta_i$	-9.368*** (0.533)	-7.103*** (0.464)
Constant	-41.901*** (2.445)	-47.444*** (2.547)
Observations	2,065	2,065
Log Likelihood	-455.661	-399.776
Akaike Inf. Crit.	927.321	813.552

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01