Moment Inequalities in Applied Work in IO*

John Asker

February 16, 2015

1 Introduction

This lecture will serve as a brief introduction to the use of moment inequalities in applied work in IO. Most of this material is based off of Pakes (2010) and Pakes et al. (forthcoming); see also Tamer (2003), Imbens and Manski (2004), Ciliberto and Tamer (2009), Romano and Shaikh (2010), Andrews and Jia (2012) (particularly on issues regarding inference).

Moment inequalities have proven useful in many recent IO papers to relax assumptions required to generate moment equalities; furthermore, in many circumstances the only information we have are bounds on the parameters of interest.[1]

2 Basic Setup

Many of the two-period models we have seen so far have followed a similar structure as follows.

Notation. Let $\pi(\cdot)$ be the profit earned in the second period, d_i and d_{-i} be the agent's and its competitors' choices, D_i be the choice set, $\mathcal{J}_i$ be the agent's information set and $\mathcal{E}[\cdot|\mathcal{J}_i]$ provide the agent's expectation conditional on that information.

Assuming that the observed set of decisions comprise a Nash equilibrium implies that:

$$C1: \quad sup_{d \in D_i} \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i] \leq \mathcal{E}[\pi(d_i = d(\mathcal{J}_i), \mathbf{d}_{-i}, \mathbf{y}_i, \theta_0)|\mathcal{J}_i],$$

where $\mathbf{y}_i$ is any variable (other than the decision variables) which affects the agent's profits, and the expectation is with respect to the joint distribution of $(\mathbf{d}_{-i}, \mathbf{y}_i)$ that summarizes the agent's beliefs on the likely realizations of those variables. For notation, let variables that the decision maker views as random be represented by boldface letters while realizations of those random variables will be represented by standard typeface.

Three points about C1 are worth noting.

1. There are *no restrictions* on the choice set; D can be continuous or discrete, multi-dimensional, etc.

2. C1 is a necessary condition for a Nash equilibrium (and for other weaker notions of equilibrium as well).

*The vast majority of this material is borrowed from Ariel Pakes' notes and Robin Lee's notes.

[1]Some recent IO papers that have utilized moment inequalities approaches include: Ho (2009), Holmes (2011), Crawford and Yurukoglu (2012), Ho, Ho and Mortimer (2012), Lee (2013), Ishii (forthcoming),...

3. Finally note that C1 is meant to be a rationality assumption in the sense of Savage (1954): i.e. the agent's choice is optimal with respect to the agent's beliefs. Most often we will assume that these beliefs are correct, but the framework is more general than that.

To check the Nash conditions, we must get an approximation of profits from *counterfactuals*; i.e. and approximation to what profits *would have been* had the agent made a choice which in fact it did not make. This, in turn, requires a model of how the agent thinks that $\mathbf{d}_{-i}$ and $\mathbf{y}_i$ are likely to change in response to a change in the agent's decision. The model for how the agent thinks $(\mathbf{y}_i, \mathbf{d}_{-i})$ are likely to respond to changes in d_i may depend on other variables, say $\mathbf{z}_i$, but those variables need to be *exogenous* in the sense that the agent thinks they will not change if the agent changes its decision. Condition 2 formalizes this assumption.

C2 : $\mathbf{d}_{-i} = d^{-i}(\mathbf{d}_i, \mathbf{z}_i)$, *and* $\mathbf{y}_i = y(\mathbf{z}_i, \mathbf{d}_i, \mathbf{d}_{-i})$, *and the distribution of* $\mathbf{z}_i$ *conditional on* $(J_i, d_i = d(\mathcal{J}_i))$ *does not depend on* d_i.

Notes on C2:

1. If the game is a simultaneous move game then $d^{-i}(d', \mathbf{z}_i) = \mathbf{d}_{-i}$ and there is no need for an explicit model of reactions by competitors.

2. In sequential problem, one needs an assumption of how one agent would respond to behavior of the other agent that is "off the equilibrium path." One will need to model or take a stance on this.

3. If the profit function has an *endogenous* r.h.s. variable $(\mathbf{y}_i)$, i.e. one which will change if d_i changes, we need a model for how it changes.

C1 and C2 together deliver the following restriction

Let $\Delta\pi(d_i, d', d_{-i}, z_i) \equiv \pi(d_i, d_{-i}, y_i) - \pi(d', d^{-i}(d', z_i), y(z_i, d', d_{-i})),$, where d' is any alternative choice in D_i. Then C1 and C2 imply that

$$\mathcal{E}[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] \geq 0, \;\; \forall \; d' \in D_i. \tag{1}$$

Equation (1) is the moment inequality delivered by the theory. To move from it to an estimation algorithm we need to specify:

1. the relationship between the expectation operator underlying the agents decisions (our $\mathcal{E}(\cdot)$) and the sample moments that the data generating process provides; and

2. the relationship between our constructed profit function and our observable measures of the determinants of those profits on the one hand, and the $\pi(\cdot, \theta)$ and (z_i, d_i, d_{-i}) that are utilized by the agent when making decisions on the other.

These two aspects of the problem differ depending on the application.

3 Relationship to Two-Period Entry/Exit Models.

The first generation two period entry/exit models that we covered used the following two assumptions.

First, the relationship between the data generating process and the agents' expectations is that:

$$FC3: \quad \forall d \in D_i, \quad \pi(d, d_{-i}, z_i, \theta_0) = \mathcal{E}[\pi(d, d_{-i}, z_i, \theta_0)|\mathcal{J}_i].$$

I.e. it is assumed that all agents know both the decisions of their competitors, and the realization of the exogenous variables that will determine profits, when they make their own decision. FC3 *rules out* asymmetric and/or incomplete information, and as a consequence, all mixed strategies (i.e. this approach implicitly restricts D_i to consist only of pure strategies)[2].

Second, there is an assumption on the relationship between the variables we measure and the variables that enter the theoretical model:

FC4. $\pi(\cdot, \theta)$ *is known.*
$z_i = (\nu_{2,i}, z_i^o)$.
$(d_i, d_{-i}, z_i^o, z_{-i}^o)$ *are observed.*
$(\nu_{2,i}, \nu_{2,-i})|_{z_i^o, z_{-i}^o} \sim F(\cdot; \theta)$, *for a known function* $F(\cdot, \theta)$.

FC4 assumes there are no errors in our profit measure; that is were we to know $(d_i, d_{-i}, z_i, z_{-i})$, we could construct an exact measure of profits for each θ. However a (possibly vector valued) component of the determinants of the profits (of the z_i) is not observed by the econometrician (denoted $\nu_{2,i}$). Since FC3 assumes full information, both $\nu_{2,i}$ and $\nu_{2,-i}$ are assumed to be known to all agents when they make their decisions, just not to the econometrician. FC4 also assumes that there is no error in the observed determinants of profits (in the z_i^o) and that the econometrician knows the distribution of $(\nu_{2,i}, \nu_{2,-i})$ conditional on (z_i^o, z_{-i}^o) up to a parameter vector to be estimated.

Substituting FC3 and FC4 into equation (1) we obtain

$$\forall d' \in D_i, \;\; \Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta_0) \;\geq 0, \tag{2}$$

and

$$(\nu_{2,i}, \nu_{2,-i})|_{z_i^o, z_{-i}^o} \sim F(\cdot; \theta_0).$$

To insure that the model assigns positive probability to the condition that

$$\forall d' \in D_i, \;\; \Delta\pi(d_i, d', d_{-i}, z_i^o, \nu_{2,i}; \theta) \;\geq 0$$

for some θ and all i (as is assumed by the model), we need further conditions on $F(\cdot)$ and/or $\pi(\cdot)$. The additional restrictions typically imposed are that the profit function is additively separable in the unobserved determinants of profits, that is that

$$\forall d \in D_i, \;\; \pi(d, d_{-i}, z_i^o, \nu_{2,i}) = \pi^{as}(d, d_{-i}, z_i^o, \theta_0) + \nu_{2,i,d}, \tag{3}$$

and that the distribution $\nu_{2,i} \equiv \{\nu_{2,i,d}\}_{d \in D_i}$, conditional on $\nu_{2,-i}$, has full support.

Some points to remember.

- The additive separability in equation (3) *cannot* be obtained definitionally by assuming ν_2 that is a residual from a projection. We could obtain the expectation of the true profit function onto variables and then get a residual, but that residual would be orthogonal to d_i and d_{-i}. However, $\nu_{2,i}$ is a determinant of the agent's decision and hence not orthogonal to d_i, and since their is full information, also not orthogonal to d_{-i}.

[2] As stated it also rules out the analysis of sequential games in which an agent who moves initially believes that the decisions of an agent who moves thereafter depends on its initial decision. However at the cost of only notational complexity we could allow for a deterministic relationship between a component of d_{-i} and (d, z_i).

- The early work on entry was looking for a useful reduced form (one that could be used to summarize the effects of environmental characteristics of the market on number of participating agents). It typically assumed orthogonality of the error and then solved out for the optimal decision of each agent (whether to enter or not to enter). They tended to find that the implied profits increased with the number of competitors. This was because more firms entered in more profitable markets (alternatively the error had components that were common to all participating agents, and hence were correlated with d_{-i}). So we needed a different way of obtaining a meaningful summary.[3]

- Say now we wanted a reduced form for our problem. To get it we could regress $\pi^{as}(\cdot)$ on variables of interest (e.g. d_{-i} and other things). Were we to do so, we would pick up an additional error which is orthogonal to these variables and which is, by construction, orthogonal to the included variables. Then we would have to deal with both errors in estimation, and they have different properties. I point this out because the next generation of models we deal with are often after such a reduced form, but they do not allow for the latter error.

The simple two-period entry model examples we covered in the previous lectures – both with homogeneous firms, and heterogeneous firms that differed in fixed costs and/or continuation values – utilized these assumptions. (NB: some papers account for asymmetric information, and thus weakened $FC3$ and $FC4$ above.)

4 Motivating Single Agent Example (Katz 2007)

Estimate the costs shoppers assign to driving to a supermarket: important to analyze zoning regulations, public transportation projects, etc. This is difficult to analyze using standard discrete choice models due to the complexity of the choice set facing consumers (all possible bundles of goods at all possible supermarkets), and assumptions needed to use standard approaches are often very strong.

Assume that an agent's utility function is additively separable in:

- utility from basket of goods bought,
- expenditure on that basket, and
- drive time to the supermarket.

Let $d_i \equiv \{b_i, s_i\}$ where b_i is the basket of goods bought, s_i is the store chosen; let z_i be individual characteristics. Then:

$$\pi(d_i, z_i, \theta) = U(b_i) - e(b_i, s_i) - \theta_i dt(s_i, z_i),$$

where $e(\cdot)$ provides expenditure, $dt(\cdot)$ provides drive time, and I have used the free normalization on expenditure (the cost of drive time are in dollars).

[3]Though when there are interacting agents the standard type of reduced form assumptions used to generate discrete choice models do not do well, there is always a reduced form for the single agent model that does make sense. Simply regress profits on variables of interest, assume a conditional distribution of the error, compute the choice as a function of the error, and form a standard estimator. The only possible problem here is a misspecification of the distribution of the errors (which can be problematic, a point we will come back to). It is the fact that this does not work for multiple agent problems that lead to the developments below.

"Profit Inequality" approach. Compare the utility from the choice the individual made to that of an alternative feasible choice. Expected difference should be positive. Requires: finding an alternative choice that allows us to isolate the effects of drive time.

For a particular d_i chose $d'(d_i)$ to be the purchase of

- the *same basket* of goods b_i,
- at a store which is *further away* from the consumer's home then the store the consumer shopped at.

Let $\mathcal{E}(\cdot)$ be the *agent's* expectation operator . Then we assume that

$$\mathcal{E}\Big[\Delta\pi(d_i, d'(d_i), z)|\mathcal{J}_i\Big] = -\mathcal{E}\Big[\Delta e(d_i, d'(d_i))|\mathcal{J}_i\Big] - \theta_i\, \mathcal{E}\Big[\Delta dt(d_i, d'(d_i))|\mathcal{J}_i\Big] \geq 0.$$

Two notes:

1. Need not specify the utility from different baskets of goods; i.e. it allows us to hold fixed the dimension of the choice that generated the problem with the size of the choice set, and investigate the impact of the dimension of interest (travel time) in isolation. Which inequalities we chose is the "sample" design question in this context. Notice also that we need not specify an outside alternative to estimate this model, whereas we typically do in discrete choice models.
2. I have not assumed that the agent's perceptions of prices are "correct" in any sense (see the discussion of the last lecture). I come back to what I need of the agent's subjective expectations.

Case 1: $\theta_i = \theta_0$. More generally all determinants of drive time are captured by variables the econometrician observes and includes in the specification. Assume that

$$N^{-1}\sum_i \mathcal{E}\Big[\Delta e(b_i, s_i, s_i')\Big] - N^{-1}\sum \Delta e(b_i, s_i, s_i') \to_P 0,$$

$$N^{-1}\sum_i \mathcal{E}\Big[\Delta dt(b_i, s_i, s_i')\Big] - N^{-1}\sum_i \Delta dt(b_i, s_i, s_i') \to_P 0$$

which would be true if, for e.g., agents were correct on average (this is stronger than we need). Then (assuming that s_i' is further away than s_i,

$$-\mathcal{E}\Big[\Delta e(d_i, d'(d_i))\Big] - \theta\, \mathcal{E}\Big[\Delta dt(d_i, d'(d_i))\Big] \geq 0$$

implies that

$$-\frac{\sum_i \Delta e(b_i, s_i, s_i')}{\sum_i \Delta dt(b_i, s_i, s_i')} \to_p \underline{\theta} \leq \theta_0.$$

(i.e., observing consumers going to closer stores provides a lower bound on travel costs).

If we would have also taken an alternative store s'' which was closer to the individual then we can form the following inequality:

$$-\frac{\sum_i \Delta e(b_i, s_i, s_i'')}{\sum_i \Delta dt(b_i, s_i, s_i'')} \to_p \overline{\theta} \geq \theta_0.$$

and we would have consistent estimates of bounds on θ_0.

Case 2: $\theta_i = (\theta_0 + \nu_{2,i})$, $\sum \nu_{2,i} = 0$. This case allows for a component of the cost of drive times ($\nu_{2,i}$) that is known to the agent (since the agent conditions on it when it makes its decision) but not to the econometrician. Then provided $dt(s_i)$ and $dt(s_i'))$ are known to the agent, then

$$\mathcal{E}\left[-\frac{\Delta e(b_i, s_i, s_i')}{\Delta dt(b_i, s_i, s_i')} - (\theta_0 + \nu_{2,i}) \right] \leq 0,$$

and provided agents expectation on expenditures are not "systematically" biased, we can get a similar inequality:

$$-\frac{1}{N} \sum_i \left(\frac{\Delta e(b_i, s_i, s_i')}{\Delta dt(b_i, s_i, s_i')} \right) \rightarrow_P \underline{\theta} \leq \theta_0.$$

(average of ratio as opposed to ratio of averages to control for heterogeneity at the individual level).

Notes.

- We did not need to specify (or compute) the utility from all different choices, so there could have been (unobserved or observed) sources of heterogeneity in the $U(b_i)$. Our choice of alternative simply differences them out.
- Case 2 allows for unobserved heterogeneity in the coefficient of interest and does not need to specify what the distribution of that unobservable is. In particular it can be *freely correlated* with the right hand side variable. "Drive time" is a choice variable, so we might expect it to be correlated with the perceived costs of that time (with ν_i).
- If the unobserved determinant of drive time costs (ν_i) is correlated with drive time (dt) then Case 1 and Case 2 estimators should be different, if not they should be the same. So there is a test for whether any unobserved differences in preferences are correlated with the "independent" variable.

Empirical Results

Data. Nielsen Homescan Panel, 2004 & data on store characteristics from Trade Dimensions. Chooses families from Massachusetts.

Discrete Choice Comparison Model. The multinomial model divides observations into expenditure classes, and then uses a typical expenditure bundle for that class to form the expenditure level (the "price index" for each outlet). Other x's are drive time, store characteristics, and individual characteristics. Note that

- the prices for the expenditure class need not reflect the prices of the goods the individual actually is interested in (so there is an error in price, and it is likely negatively correlated with price itself.)
- it assumes that the agents new the goods available in the store and their prices exactly when they decided which store to chose (i.e. it does not allow for expectational error)
- it does not allow for unobserved heterogeneity in the effects of drive time. We could allow for a random coefficient on drive time, but, then we would need a conditional distribution for the drive time coefficient....

Focus. Allows drive time coefficient to vary with household characteristics. Focus is on the average of the drive time coefficient for the median characteristics (about forty coefficients; chain dummies, outlet size, employees, amenities...).

Multinomial Model: median cost of drive time was \$240 (when the median wage in this region is \$17). Also several coefficients have the "wrong" sign or order (nearness to a subway stop, several amenities, and chain dummies).

Inequality estimators. Uses a lot of moments: point estimates, but tests indicated that the model was accepted. Standard errors are very conservative.

- Inequality estimates with

$$\theta_i = \theta_0 : \quad .204 \;\; [.126, .255]. \quad \Rightarrow \$4/hour,$$

- Inequality estimates with

$$\theta_i = \theta_0 + \nu_i : \quad .544 \; [.257, .666], \quad \Rightarrow \$14/hour$$

and other coefficients straighten out.

Apparently the unobserved component of the coefficient of drive time is negatively correlated with observed drive time differences.

5 The Revealed Preference Approach, More Generally.

We want to extend these ideas to a general choice set and allow for interacting agents. Recall that we have leverage C1 (individual rationality) and C2 (exogeneity of z') to obtain the following:

$$\mathcal{E}[\Delta\pi(d_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] \geq 0, \;\; \forall \; d' \in D_i.$$

To take this as an empirical framework, we need to formalize:

- The relationship of the expectation operator the agent uses, and the data generating process;
- The relationship between the profit function the agent perceives and the variables that determine its value, and the profit function that we (the econometricians) specify and the variables which determine its value.

We discuss each in turn.

5.1 The agent's expectation operator and the DGP.

We begin with the assumption on the relationship between the expectation operator underlying agents' decisions (our $\mathcal{E}(\cdot)$), and the expectation conditional on the process actually generating the data (our $E(\cdot)$). This will help clarify the sense in which the model can be misspecified without invalidating the properties of the estimator. Throughout we shall assume that we know a subset of the variables that are in the agent's information set when it makes its decision.

We will assume that there is a known subset of the observed variables, denoted x_i, which are contained in $\mathcal{J}_i$ and satisfy the condition that if $h(\cdot)$ is a positive valued function, then

$$PC3: \;\; \frac{1}{N}\sum_i \mathcal{E}\Big(\Delta\pi(d_i, d', d_{-i}, z_i)|x_i\Big) \geq 0 \;\; \Rightarrow \;\; E\frac{1}{N}\sum_i \Big(\Delta\pi(d_i, d', d_{-i}, z_i)h(x_i)\Big) \geq 0 \quad .$$

- Correct Expectations are Sufficient:
  - Standard Bayes-Nash condition. Each agent knows: (i) the other agents' strategies [$\mathbf{d}_{-i}(\mathcal{J}_{-i})$], and (ii) the joint distribution of other agents' information sets and the primitives sources of uncertainty [of $(\mathcal{J}_{-i}, \mathbf{z}_i)$)] conditional on $\mathcal{J}_i$, and regularity conditions.
  - Weaker condition: agents' conditional expectations of profit difference are *correct*. Does not require knowledge of other agent's strategies, or the distribution of $(\mathbf{d}_{-i}, \mathbf{z}_i)$ conditional on $\mathcal{J}_i$. E.g. auctions requires only knowledge of the distribution of second order statistic – which could be learned from experience if the same situation was repeated many times.
- Incorrect expectations are also possible; all we need is that the *average* of

$$\mathcal{E}[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i] - E[\Delta\pi(d_i, d', d_{-i}, z_i)|x_i],$$

  is non-negative. Relevant cases:
  - Agents beliefs are not exactly right but the difference between agents' expectations on $\Delta\pi_i(\cdot, \theta_0)$ and the expectation of the data generating process are mean zero conditional on x, or
  - Agents can be consistently "overly optimistic" about the relative profits from the decisions they make.

5.2 The Agent's and Econometrician's Measure of Profits

We introduce a general model of how profits are measured before providing the restrictions we will leverage. This formality makes explicit the sources of the disturbances in our models and their relationship to different estimation strategies.

Denote by $r(d, d_{-i}, z_i^o, \theta_0)$ our *observable* approximation to $\pi(\cdot)$. Then we can always define

$$\nu(d, d_{-i}, z_i^o, z_i, \theta_0) \equiv r(d, d_{-i}, z_i^o, \theta_0) - \pi(d, d_{-i}, z_i),$$

so that

$$r(\cdot) = \pi(\cdot) + \nu, \qquad \text{and} \qquad \mathcal{E}[r(\cdot)|\cdot] = \mathcal{E}[\pi(\cdot)|\cdot] + \mathcal{E}[\nu|\cdot].$$

Then

$$r(d, d_{-i}, z_i^o, \theta_0) \equiv \mathcal{E}[\pi(d, \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] + \nu_{2,i,d} + \nu_{1,i,d},$$

where

$$\nu_{2,i,d} \equiv \mathcal{E}[\nu(d, \mathbf{d}_{-i}, \mathbf{z}_i^o, \mathbf{z}_i, \theta_0)|\mathcal{J}_i],$$

$$\nu_{1,i,d} \equiv \Big(\pi(d, \cdot) - \mathcal{E}[\pi(d, \cdot)|\mathcal{J}_i]\Big) + \Big(\nu(d, \cdot) - \mathcal{E}[\nu(d, \cdot)|\mathcal{J}_i]\Big) \equiv \nu_{1,i,d}^{\pi} + \nu_{1,i,d}^{r}.$$

Sources of ν_1. The first source of differences between our approximation to profits and an agents expectations of profits will be comprised of: *expectational error* from incomplete (uncertainty in $\mathbf{z}_i$), and/or asymmetric (uncertainty in $\mathbf{d}_{-i}$) information,

$$\nu_1^{\pi} = \pi(d, \cdot) - \mathcal{E}[\pi(d, \cdot)|\mathcal{J}_i]$$

and *specification and measurement error* or

$$\nu_1^r = \nu(d, \cdot) - \mathcal{E}[\nu(d, \cdot)|\mathcal{J}_i]$$

(this includes the error from projecting profits onto variables on interest in reduced form models).

Note that:

- Although $\mathcal{E}[\nu_{1,i,d}|\mathcal{J}_i] = 0$, by construction, it will often be the case that $\mathcal{E}[\nu_{2,i,d}|\mathcal{J}_i] \neq 0$. This distinction is why we need to keep track of two separate disturbances.

- To compute the distribution of the expectational error we would have to specify what each agent knows about its competitors, and then repeatedly solve for an equilibrium (a process which typically would require us to select among equilibria).

ν_2 and Selection. ν_2 is that part of profits that the agent conditions on when making its decision but is not included in the econometrician'sprofit specification. Since $\nu_{2,i} \in \mathcal{J}_i$ and $d_i = d(\mathcal{J}_i)$, d_i will generally be a function of $\nu_{2,i}$ (and perhaps also of $\nu_{2,-i}$). This can generate a selection problem.

To see why, temporarily assume that the agent's expectations (our $\mathcal{E}(\cdot)$) equals the expectations generated by the true data generating process (our $E(\cdot)$), that x is an "instrument" in the sense that $\mathcal{E}[\nu_2|x] = 0$, and that $x \in \mathcal{J}$. Then

$$\mathcal{E}[\nu_1|x] = \mathcal{E}[\nu_2|x] = 0.$$

These expectations *do not* condition on d_i, and any moment which depends on d_i requires properties of the disturbance conditional on d_i. Since d is measurable $\sigma(\mathcal{J})$

$$\mathcal{E}[\nu_1|x, d] = 0.$$

However since $\nu_2 \in \mathcal{J}$ and

$$\mathcal{E}[\pi(\cdot)|\cdot] = \mathcal{E}[r(\cdot)|\cdot] + \nu_2,$$

if the agent chooses d^* then

$$\nu_{2,d^*} - \nu_{2,d} \geq \mathcal{E}[r(\cdot, d)|\cdot] - \mathcal{E}[r(\cdot, d^*)|\cdot]$$

so

$$\mathcal{E}[\nu_{2,d^*}|x, d^*] \neq 0, \quad and \quad \mathcal{E}[\nu_{2,d}|x, d] \neq 0.$$

The fact that "x is an instrument" does not "solve" the selection problem. In particular, in our inequality context an estimation algorithm based on accepting any value for θ which makes the sample average of our observable proxy for the difference in profits (of $\Delta r(\cdot, \theta)$), or its covariance with a positive valued instrument, positive *should not*, at least in general, be expected to lead to an estimated set which includes θ_0 (even asymptotically).

5.3 Overcoming the Selection Problem.

The econometrician only has access to $\Delta r(\cdot, \theta)$ and our best response condition is in terms of the conditional expectation of $\Delta\pi(\cdot)$. So we need an assumption which enables us to restrict weighted averages of $\Delta r(\cdot)$ in a way that insures that the expectation of the weighted average of $\Delta r(\cdot, \theta)$ is positive at $\theta = \theta_0$.Here are two ways around it that are frequently used.

PC4a: Differencing. Here there are groups of observations with the same value for the ν_2 error. We end up getting difference in difference inequalities (the difference for one observation contains the same ν_2 error as the difference for the other).

Let there be G groups of observations indexed by g, counterfactuals $d'_{i,g} \in \mathcal{D}_{i,g}$, and positive weights $w_{i,g} \in \mathcal{J}_{i,g}$, such that $\sum_{i \in g} w_{i,g} \Delta\nu_{2,i,g,d_{i,g},d'_{i,g}} = 0$; i.e. a within-group weighted average of profit differences eliminates the ν_2 errors. Then

$$G^{-1} \sum_g \sum_{i \in g} w_{i,g} \Big(\Delta r(d_{i,g}, d'_{i,g}, \cdot; \theta_0) - \mathcal{E}[\Delta\pi(d_{i,g}, d'_{i,g}, \cdot; \theta_0) | \mathcal{J}_{i,g}] \Big) \to_P 0,$$

provided $G^{-1} \sum_g \sum_i w_{i,g} \Delta r(d_{i,g}, d'_{i,g}, \cdot; \theta_0)$ obeys a law of large numbers.

Our supermarket example is a special case of $PC4a$ with $n_g = w_{i,g} = 1$. There $d_i = (b_i, s_i)$, $\pi(\cdot) = U(b_i, z_i) - e(b_i, s_i) - \theta_0 dt(s_i, z_i)$ and $\nu_{2,i,d} \equiv U(b_i, z_i)$. If we measure expenditures up to a $\nu_{1,i,d}$ error, $r(\cdot) = -e(b_i, s_i) - \theta_0 dt(s_i, z_i) + \nu_{2,i,d} + \nu_{1,i,d}$. We chose a counterfactual with $b'_i = b_i$, so $\Delta r(\cdot) = \Delta\pi(\cdot) + \Delta\nu_{1,\cdot}$, and the utility from the bundle of good bought is differenced out. "Matching estimators", i.e. estimators based on differences in outcomes of matched observations, implicitly assume $PC4a$ (no differences in unobservable determinants of the choices made by matched observations).

PC4b: Unconditional Averages and Instrumental Variables. PC4b assumes there is a counterfactual which gives us an inequality that is additive in ν_2 *no matter the decision the agent made.* The counterfactual may be different for different observations. Then we can form averages *which do not condition on* d, and hence do not have a selection problems.

Assume that $\forall d \in D_i$, there is a $d' \in D_i$ and a $w_i \in \mathcal{J}_i$ such that

$$w_i \Delta r(d_i, d'_i, \cdot; \theta) = w_i \mathcal{E}[\Delta\pi(d_i, d'_i, \cdot; \theta) | \mathcal{J}_i] + \nu_{2,i} + \Delta\nu_{1,i,\cdot},$$

Then if $x_i \in \mathcal{J}_i$, $E[\nu_{2,i} | x_i] = 0$, and $h(\cdot) > 0$

$$N^{-1} \sum_i w_i \Delta r(d_i, d'_i, \cdot; \theta_0) h(x_i) \ \to_P N^{-1} \sum_i w_i \mathcal{E}[\Delta\pi(d_i, d'_i, \cdot; \theta_0) | \mathcal{J}_i] h(x_i) \geq 0,$$

provided $N^{-1} \sum \nu_{1,i,\cdot} h(x_i)$ and $N^{-1} \sum \nu_{2,i} h(x_i)$ obey laws of large numbers.

Case 2 of our supermarket example had two ν_2 components; a decision specific utility from the goods bought, $\nu_{2,i,d} = U(b_i, z_i)$ (like in case 1), and an agent specific aversion to drive time, $\theta_i = \theta_0 + \nu_{2,i}$. As in case 1, taking $d' = (b_i, s'_i)$ differenced out the $U(b_i, z_i)$. Then $\Delta r(\cdot) = -\Delta e(\cdot, s_i, s'_i) - (\theta_0 + \nu_{2,i}) \Delta dt(s_i, s'_i, z_i) + \Delta\nu_{1,\cdot}$. Divide by $\Delta dt(s_i, s'_i, z_i) \leq 0$. Then $C1$ and $C2$ imply that $\mathcal{E}[\Delta e(s_i, s'_i, b_i) / \Delta dt(s_i, s'_i, z_i) | \mathcal{J}_i] - (\theta_0 + \nu_{2,i}) \leq 0$. This inequality is; (i) linear in $\nu_{2,i}$, and (ii) is available for every agent. So if $E[\nu_2] = 0$, $PC3$ and a law of large numbers insures $N^{-1} \sum_i \nu_{2,i} \to_P 0$, and $\sum_i \Delta e(s_i, s'_i, b_i) / \Delta dt(s_i, s'_i, z_i) \to_P \underline{\theta}_0 \leq \theta_0$; while if $E[\nu_2 | x] = 0$ we can use x to form instruments. Notice that $\nu_{2,i}$ can be correlated with $dt(z_i, s_i)$ so this procedure enables us to analyze discrete choice models when a random coefficient affecting tastes for a characteristic is correlated with the characteristics chosen.

5.4 Left For the Reader: Formalities on What is Needed and Some Examples.

By appropriate choice of weighting functions we can do quite a bit. We consider functions of the form

$$h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i}) : \mathcal{D}_i \to \mathcal{R}^+$$

So $h(\cdot)$ has to be non-negative and its value can depend on;
(i) the alternative choice considered (on d'),
(ii) on variables in the information set $\mathcal{J}_i$ (which determines d_i),
(iii) and (possibly) on some observable component of the other agents' information sets, $\mathbf{x}_{-i} \subset \times_{j\neq i}\mathcal{J}_j$.

Notice that by allowing weights to depend on $\mathbf{x}_{-i}$ we no longer can insure that the weights are mean independent of ν_1 as we only insured that $\mathcal{E}\Delta r(\cdot)$ was independent of $\mathcal{J}_i$ and $\mathbf{x}_{-i} \notin \mathcal{J}_i$.

Definitions For $i = 1, ..., n$ and $(d, d') \in \mathcal{D}_i^2$, define:

$$\begin{aligned}
\nu_{2,i,d,d'} &= \mathcal{E}[\Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i)|\mathcal{J}_i] - \mathcal{E}[\Delta r(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i^o)|\mathcal{J}_i] \\
\nu_{1,i,d,d'} &= \nu^{\pi}_{1,i,d,d'} - \nu^{r}_{1,i,d,d'} \\
\nu^{\pi}_{1,i,d,d'} &= \Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i) - \mathcal{E}[\Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i)|\mathcal{J}_i] \\
\nu^{r}_{1,i,d,d'} &= \Delta r(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i^o) - \mathcal{E}[\Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i^o)|\mathcal{J}_i]
\end{aligned}$$

From these definitions, it follows that:

$$\Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i) = \Delta r(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i^o, \theta_0) + \nu_{1,i,d,d'} + \nu_{2,i,d,d'}$$

and, re-arranging terms and taking expectations (note $E[\nu^{\pi}_{1,i,d,d'}|\mathcal{J}_i] = 0$) yields:

$$\Delta r(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i^o, \theta_0) = E[\Delta\pi(d, d', \boldsymbol{d}_{-i}, \boldsymbol{z}_i) + \nu^{r}_{1,i,d,d'} - \nu_{2,i,d,d'}$$

What is needed? Given these weighting function our assumption is designed to insure that

$$E\Big[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})\Delta r(d', \mathbf{d}_i, \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)\Big]$$

Since

$$\Delta r(d, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0) = E[\Delta\pi(d, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] + \nu^{r}_{1,i,d,d'} - \nu_{2,i,d,d'}$$

the inequality we want to take to data is

$$\begin{aligned}
& E\left[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})\Delta r(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i^o, \theta_0)\right] \\
= \ & E\left[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})E[\Delta\pi(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i]\right] \qquad (4) \\
& + E\left[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})\nu^{r}_{1,i,\mathbf{d}_i,d'}\right] - E\left[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i} h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})\nu_{2,i,\mathbf{d}_i,d'}\right]
\end{aligned}$$

Since $E[\Delta\pi(\mathbf{d}_i, d', \mathbf{d}_{-i}, \mathbf{z}_i)|\mathcal{J}_i] \geq 0$ each term in the first summand is nonnegative by the assumed nonnegativity of the weights $h^i(d'; \mathbf{d}_i, \mathcal{J}_i, \mathbf{x}_{-i})$. However we also require the last two summations to be non-negative.

As noted the definition of ν_1^r in equation yields $E[\nu^r_{1,i,\mathbf{d}_i,d'}|\mathcal{J}_i]=0$. So, when the weight function does not depend on $\mathbf{x}_{-i}$, the summation over ν_1^r terms in equation (4) is zero. However when $h^i(\cdot)$ depends non-trivially on the actions or information sets of other agents, then ν_1^r can be correlated with $h^i(\cdot)$ and this will violate the condition we now turn to.

If we can insure that the last term is negative, then C2 holding will imply that this inequality holds

Condition PC4 Let $h^i(d';\mathbf{d}_i,\mathcal{J}_i,\mathbf{x}_{-i})$ be a nonnegative function whose value can depend on the alternative choice considered (on d'), on the information set $\mathcal{J}_i$ (which determines d_i), and (possibly) on some observable component of the other agents' information sets, $\mathbf{x}_{-i}\subset\times_{j\neq i}\mathcal{J}_j$. Assume that

$$E[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i}h^i(d';\mathbf{d}_i,\mathcal{J}_i,\mathbf{x}_{-i})\Delta\nu_{2,i,\mathbf{d}_i,d'}]\leq 0$$

and

$$E[\sum_{i=1}^{n}\sum_{d'\in\mathcal{D}_i}h^i(d';\mathbf{d}_i,\mathcal{J}_i,\mathbf{x}_{-i})\Delta\nu^r_{1,i,\mathbf{d}_i,d'}\geq 0. \spadesuit$$

To see what underlies $PC4$ assume temporarily that for each d_i, $h^i(d';\mathbf{d_i},\mathcal{J}_i)\equiv 1$ for a particular $d'(d_i)$ and 0 elsewhere. Then it suffices that an average over agents of

$$\sum_{j\in D_i}E[\big(\nu_{2,j}-\nu_{2,d'(j)}\big)|\,d_i=j,x_i]\,\Pr\{d_i=j|x_i\}$$

is not positive. This is an *unconditional* average, it sums over each possible choice for each individual (the choice actually made enters in the same way as any other possibility). For each choice the researcher is free to chose any alternative provided the average of the *differences* between the ν_2 associated with the choice and its alternatives is not positive.

The function $h(\cdot)$ allows the researcher to weight the different differences differently. Consider the supermarket example in which the cost of drive time differed across shoppers. We first chose an alternative which was buying the same bundle at different stores thus differencing out the impact of heterogeneity in preferences over bundles (i.e. we set all other $h(\cdot)=0$), then we weighted the resultant utility differences across individuals by the inverse of the difference in drive time thus isolating the drive time coefficients (so for that $d'(j)$, $h(j,d'(j),x)=1/\Delta dt(j,d'(j)))$. We only then averaged over individuals to obtain the average cost of drive time.

5.4.1 Assumptions which imply $PC4$.

The flexibility in this approach comes from the ability to chose the function $d'(d_i)$; as this can be chosen to be any decision in the agent's choice set. We try to chose it to enable us to account for unobservables that agent's knew prior to the choice but are not contained in our data sets. Here is an outline of ways that have been found to do so.

Cases when we can form inequalities which do not depend on the models' $\nu_{2,i,d}$ by appropriate choice of $h(j,d(j),\cdot)$.

- $\forall d,\nu_{2,i,d}=\nu_{2,i}$, that is when the unobservable known to the agent when it makes its choice but not observed by the econometrician is constant across choices, or

- $\nu_{2,i,d}$ can vary across decisions, but the same value of $\nu_{2,i,d}$ appears in more than one of them (so there are "group effects").

This often allows us to use perturbation conditions, similar to those used by Euler for continuous choice problems, to analyze problem with discreteness in choice sets, or at boundaries of the choice set. For one example see Morales, 2011.

- models for micro data where a variable needed for an inequality is unobserved (or is measured with error) at the micro level but is observed (or the error is averaged out) at a higher level of aggregation (say because of the availability of Census data).

Cases where we can form inequalities that do not condition on $\nu_{2,i,d}$ being in a particular set.

- Ordered choice models (defined broadly enough to include the vertically differentiated demand model used in I.O.) satisfy $PC4$.
- contracting models in which there is a component of the contract that the agents know but the econometrician does not observe (as in Ho 2009, which we will cover later.)

These are cases which appears often in IO models because the unobservable is often a determinant of cost that is unknown to the econometrician but known to the agents. Here are two familiar cases in point.

- The firm is buying a discrete number of units, so $d_i \in \mathcal{Z}_+$. We can always take as $d'(j) = j+1$ and the difference in profits will contain the cost savings from not purchasing the additional unit (in this case are instrument has to be orthogonal to the unobserved determinant of cost).
- Contracts between buyers and sellers which have an unobserved component, which is, say, a cost to the buyer. Then the cost is a profit to the seller if the contract is established, and is a savings to the buyer if the contract is not accepted.

Functionals Form Assumption. None of the cases above required assumptions on the functional form of the disturbances. Sometimes when we can not get rid of the selection problem without some assumption, an assumption on the form of the distribution of ν_2 will be enough. Below we note a case where all we need is symmetry of the ν_2 errors will enable one to maintain the inequality by using the tail of one side to correct for selection on the other (this in the spirit of Powell's (1986) symmetry assumption). Several cases with full functional form assumptions are now being developed.

This is a new literature, and there are different ways of dealing with the selection problem currently being developed. Most of them is add restrictions to the distribution of disturbances; but still do not require a particular distribution for them.

6 Discrete Investment Choices by A Firm.

This application is due to Ishii 2008. It is about analyzing choices of a number of ATM's but as will become obvious similar analysis could be used for at least some types of entry games.

Ishii analyzes how ATM networks affect market outcomes in the banking industry. The part of her study we consider here is the choice of the number of ATMs. More generally these papers provide techniques that can be used to empirically analyze "lumpy" investment decisions, or investment decisions subject to adjustment costs which are not convex for some other reason[4], in market environments. Ishii uses a two period model with simultaneous moves in each period.

- First period; each bank chooses a number of ATMs to maximize its expected profits given its perceptions on the number of ATMs likely to be chosen by its competitors.
- Second period interest rates are set conditional on the ATM networks in existence and consumers chose banks.

Note that there are likely to be many possible Nash equilibria to this game.

How do we get to second stage profit function? Initial stages.

- estimate a demand system for banking services and an interest rate setting equation; both conditional on the number of ATMs of the bank and its competitors, i.e. on (d_i, d_{-i}).

Discrete choice model among a finite set of banks with consumer and bank specific unobservables (as in BLP). The indirect utility of the consumer depends on the distance between the consumer's home and the nearest bank branches, the consumer's income, interest rates on deposits, bank level of service proxies, the size of the ATM network, and the distribution of ATM surcharges (surcharges are fees that consumers pay to an ATM owner when that owner is not the consumer's bank).

- Interest rates are set in a simultaneous move Nash game. Note: the model is structural so we can also compute what interest rates would be and where consumers would go were their a different network of ATMs.

Note. Need to assume that the solution to the second stage is unique; or at least that you are calculating the one all participants agree would occur. Come back to the realism of this below.

Compute the banks' earnings conditional on the ATM networks in existence, say $r(y_i, d_i, d_{-i})$. These are calculated as the earnings from the credit instruments funded by the deposits minus the costs of the deposits (including interest costs) plus the fees associated with ATM transactions. The ATM fee revenue is generated when non-customers use a bank's ATMs and revenue is both generated and paid out when customers use a rival's ATMs.

The ATM Choice Model. To complete the analysis of ATM networks Ishii requires estimates of the cost of setting up and running ATMs. Crucial to the analysis of the implications of existing network (is there over or under investment, are ATM networks allowing for excessive concentration and excessively low interest on customer accounts,...) and of what the network is likely to result from alternative institutional rules (of particular interest is the analysis of systems that do not allow surcharges, as suggestions to eliminate surcharges have been part of the public debate for some time).

We infer what cost must have been for the network actually chosen to be optimal. So we model choice network size; of $d_i \in \mathcal{D} \subset \mathcal{Z}^+$, the non-negative integers. We assume a simultaneous move gain. The agent forms a perception on the distribution of actions of its competitors and of likely values of the variables that determine profits in the next period, and chooses the d_i that maximizes expected profits. So this is a multiple agent ordered choice model. Formally

$$\mathcal{E}[\pi(y_i, d_i, d_{-i}, \theta)|\mathcal{J}_i] = \mathcal{E}[r(y_i, d_i, d_{-i})|\mathcal{J}_i] - (\theta + \nu_{2,i})d_i, \tag{5}$$

where

[4] Actually Ishii's problem has two sources of non-convexities. One stems from the discrete nature of the number of ATM choice, the other from the fact that network effects can generate increasing returns to increasing numbers of ATMs

- $\mathcal{J}_i$ is the information known by the agents when the decisions on the number of ATM's must be made,
- θ is average cost of an ATM, and the ν_i capture the effects of cost differences among banks that are unobserved to the econometrician but known to the agent. What we know is there are a set of instruments such that $E[\nu_{2,i}|x_i] = 0$

Clearly a necessary condition for an optimal choice of d_i is that:

- expected profits from the observed d_i is greater than the expected profits from $d_i - 1$
- expected profits from the observed d_i is greater than the expected profits from $d_i + 1$.

Since we can calculate what the bank would earn in income in both those situations, these two differences provide inequalities that the costs of ATMs must satisfy, and when we average them over banks, they provide an inequality estimator of θ.[5]

The inequality for the first case is[6]

$$0 \le \mathcal{E}[\pi(y_i, d_i, d_{-i}, \theta)|\mathcal{J}_i] - \mathcal{E}[\pi(y_i, d_i - 1, d_{-i}, \theta)|\mathcal{J}_i] =$$
$$\mathcal{E}[r(y_i, d_i, d_{-i})|\mathcal{J}_i] - \mathcal{E}[r(y_i, d_i - 1, d_{-i})|\mathcal{J}_i] - (\theta + \nu_{2,i})$$

This will give us an upper bound for θ. I will let you work out the second case. It gives us our lower bound.

A few points are worthy of note.

- Note we have chosen $d'(d_i)$ in a way that insures we keep a $\nu_{2,i}$ for every agent (there is no selection).
- To do this we need to solve out for the returns that would be earned were there a different ATM network (for $r(y_i, d_i - 1, d_{-i})$, etc.) $\Rightarrow$ we have to solve out for the interest rates that would prevail were the alternative networks chosen. This is why you need the structural static model; i.e. we need approximations to counterfactuals.
- The expectation is conditional on information known when the decisions are made. It is over any component of y_i not known at the time decisions are made, and over the actions of the competitors (over d_{-i}). Note that we do not need to specify what that information set is.

Our behavioral assumptions imply.

$$E\left(r(y_i, d_i, d_{-i}) - r(y_i, d_i - 1, d_{-i}) - (\theta_0 + \nu_{2,i}) \right) \ge 0$$

[5] These conditions will also be sufficient if the expectation of $\pi(\cdot)$ is (the discrete analogue of) concave in d_i for all values of d_{-i}, a condition which works out to be almost always satisfied at the estimated value of θ.

[6] More formally to get this we use $PC4$ substituting

$$h\big(j, d'(j), \cdot\big) = 1 \;\; if j = d_i; \;\; h\big(j, d'(j), \cdot\big) = -1 \;\; if j = d_i - 1,$$

and $h\big(j, d'(j), \cdot\big) = 0$ elsewhere.

and

$$E\Bigg(r(y_i, d_i, d_{-i}) - r(y_i, d_i + 1, d_{-i}) + (\theta_0 + \nu_{2,i})\Bigg) \leq 0,$$

with $\sum \nu_{2,i} = 0$ by construction. If we had an instrument (an x which is the in the agents' information set when it made its decision) that was orthogonal to $\nu_{2,i}$ and $h(\cdot)$ was a positive value function, our behavioral assumptions would also imply

$$\sum_i E\Bigg(r(y_i, d_i, d_{-i}) - r(y_i, d_i - 1, d_{-i}) - (\theta_0 + \nu_{2,i})\Bigg)h(x_i) \geq 0$$

Simplest Estimator. Let $\Delta\overline{r}_L$ be the sample average of the returns made from the last ATM installed, and $\Delta\overline{r}_R$ be the sample average of the returns that would have been made if one more ATM had been installed. Then

$$\Delta\overline{r}_L - \theta \geq 0 \quad (i.e.\ \ \Delta\overline{r}_L \geq \theta),$$

and

$$-\Delta\overline{r}_R + \theta \geq 0 \quad (i.e.\ \ \theta \geq \overline{r}_R).$$

Assuming $\Delta\overline{r}_R \leq \Delta\overline{r}_L$

$$\hat{\Theta}_J = \{\theta :\ -\Delta\overline{r}_R \leq \theta \leq \Delta\overline{r}_L\}.$$

Notes. With more instruments the *lower bound* for θ_0 is the *maximum* of a finite number of moments, each of which distribute (approximately) normally. So actual lower bound has a positive bias in finite samples. The estimate of the upper bound is a minimum, so the estimate will have a negative bias. $\Rightarrow \hat{\Theta}_J$ may well be a point even if Θ_0 is an interval. Importance of test.

Results (see table).

- $h(x) =$ constant $\Rightarrow$ interval, $h(x) =$ all $\Rightarrow$ a point.

  (Instruments: transformation of market population, the number of banks in a market, and the number of branches of a bank in a market (indicator if each value is above or below mean). These variables are all in the information sets of the agents at the time they make their ATM network choices).

- Test: $d_i \notin$ IV accepts, $d_i \in IV$ rejects.

- CI pretty tight, and pretty stable across specifications ($\approx$ \$4,500 per ATM per month). Quite a bit larger than prior estimates which do not take into account all aspects of costs.

Implications. Ishii (forthcoming). Large banks subsidize their ATM networks in order to gain customers (whom they pay lower interest rates to). The question of whether to force equal access to all ATMs and a central surcharge was considered in congress. She considers a counterfactual with the same number of ATMs, imposes a universal ATM user fee that would just cover ATM costs, and recalculates equilibrium. A centralized surcharge would reallocate profits from large to small banks and decrease concentration markedly. Welfare effects (conditional on the network) not as obvious because of costs of ATMs. She also show that investment in ATMs is suboptimal; so one might want to make the ATMs endogenous and see what happens, but then we get faced with, among other things, the issue of multiplicity of equilibria.

Note. We have ignored the fact that often one side of an ordered choice model is truncated. In Ishii's model some banks actually had zero ATM's and for these banks there is no feasible change to the left. If we simply ignore those banks we induce a selection problem; they may well have been banks with particularly negative cost shocks. We can treat this in at least two ways; i) find a variable which you know is greater than the value of their ATM costs, and use it to bind, or ii) assume the cost distribution is symmetric and use the symmetry to bound one side (See PPHI).

Table 1: Inequality Method, ATM Costs*

	θ_J	95% CI for θ LB	UB
$h(x) \equiv 1$	[32,006, 32,492]	23,301	41,197
$h(x) =$ full*	32,492	29,431	38,444
Different Choices of $D(d_i)$ ($h(x) =$ full)			
3.$\{d : \|d - d_i\| = 2\}$	36,188	31,560	38,947
4.$\{d : \|d - d_i\| = 1, 2\}$	36,188	31,983	36,869
Extending the Model ($h(x)$ =full)			
5.* θ_b	36,649	32,283	38,871
6.* θ_r	38,348	26,179	47,292
Test Statistics		$d \notin$ IV	$d \in$ IV
T(observed)/T(critical at 5%)		.96	1.36

There are 291 banks in 10 markets. * θ_b is in-branch cost, θ_r is remote cost.

6.1 Digression: Multiple Equilibria and Counterfactual's in Ishii's game.

Ishii's counterfactuals held fixed network. How to think about equilibrium network response?

This is taken from Lee and Pakes (2009, *Economic Letters*). Take Ishii's information on Pittsfield, Massachusetts and analyze the likely impact of a change in Pittsfield's banking environment (a hypothetical merger and unexpected shock to Pittsfield's economy which changes the costs of running an ATM).

There were eight banks before the merger, so we examine the actions of the seven remaining banks in the market. We assume the merged bank has a profit function which consists of the sum of the profits from the two banks which merged and starts with their ATMs, giving us an initial allocation of ATMs to the seven banks of $(9, 0, 3, 1, 0, 0, 1)$. Note that, as is often the case in empirical work, there is significant heterogeneity across the firms inherited from past actions and events (the banks differ in the number and locations of their branches, in the amenities they provide customers...). We are assuming that these characteristics of the banks *do not* change.

The realized costs of agent i if it uses n_i ATMs in period t are given by:

$$C(n_i, t) = [b_{0,i} + b_{1,i,t}]n_i + b_2 n_i^2$$

where $(b_{0,i}, b_2)$ are known constants and $b_{1,i,t}$ is the random draw on the cost shock. These are iid draws from a normal distribution with mean μ and variance σ^2 that is common across firms. For simplicity, we assume switching costs and fixed costs of each machine to be 0; we only focus on the per-period operational costs.

Firms do not know their future cost shocks before they chose the number of ATMs they operate in the next period, and we focus on Nash Equilibria in expected costs. In the first period after the merger, each firm receives its own realization of the cost shock $b_{1,i,t}$. As firms realize that their

Table 1: **Possible Equilibria for Four Mean Cost Specifications**

	Mean Cost (μ)	20,000	15,000	10,000	0
ATM Allocation	# of ATMs	Is Allocation An Equilibrium?			
(4,0,4,0,0,1,1)	10	Yes	No	No	No
(5,0,3,0,0,1,1)	10	Yes	No	No	No
(4,0,4,0,0,1,2)	11	No	Yes	No	No
(4,0,4,0,1,1,1)	11	No	Yes	Yes	No
(5,0,3,0,1,1,2)	12	Yes	Yes	Yes	Yes

costs have changed, each firm will use an average over cost draws after the switch in regimes to form their expectation of costs for the next period (μ). There are no dynamics other than that induced by learning about the likely value of the cost shocks and the likely play of competing firms.

6.1.1 Number and Nature of Equilibria

The first part of the analysis proceeds by simply enumerating the "limiting equilibria": i.e., the Nash equilibria when all firms know the expected value of the cost shock. Since banks are asymmetric, there are $170,544$ different allocations of up to 15 ATMs among seven banks. Table 1 lists all equilibrium allocations when firms know the expected value of the cost shock for different values of μ.

Results.

- initial post merger allocation is (9,0,3,1,0,0,1) does not constitute a best response for any of our cost specifications.
- the number of equilibria is always *strikingly small* in comparison to the number of total possible allocations.
- within a specification for costs, the different equilibria are quite similar to each other (no two equilibria for the same cost specification in which one firm differs in its number of ATMs by more than one ATM,...)
- "comparative statics"; if an allocation which had been an equilibrium is no longer an equilibrium when we lower the cost, this former equilibrium was always the equilibrium with the least number of ATMs at the higher cost. If an allocation becomes an equilibrium allocation when it had not been one at the higher cost, the new equilibrium allocation always has a larger total number of ATMs then the equilibria that are dropped out (and those that are dropped are always the equilibria with the lowest number of ATMs).

6.1.2 Equilibrium Selection through Belief Formulation.

Investigate the implications of different processes for forming beliefs about competitors' play.

- Best response; each firm believes its competitors' will play the same strategy in the current period as they did in the prior period

Table 2: **Fraction of Rest Points at Alternative Equilibria**

Mean (μ)	20,000			15,000			10,000			0		
CV (σ/μ)[a]	1	.5	.25	1	.5	.25	1	.5	.25	1	.5	.25
Best Reply												
4040011	.89	.87	.82									
5030011	.10	.10	.13									
4040012				.27	.14	.01						
4040111				.40	.21	.02	.04[b]	.00	.00			
5030112	.01	.03	.06	.33	.65	.97	.94	1.0	1.0	1.0	1.0	1.0
Fictitious Play.												
4040011	.47	.41	.41									
5030011	.34	.44	.30									
4040012				.00	.00	.00						
4040111				.10	.01	.00	.00	.00	.00			
5030112	.15	.15	.29	.90	.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0

The initial condition is (9,0,3,1,0,0,1) for all runs and is never an equilibrium based on true expected costs.
[a] CV is the coefficient of variation of the cost shock. For the base specification where $\mu = 0$, the variance of the cost shocks were set to be the same as when $\mu = 20,000$.
[b] In this specification under Best Reply, approximately 2% of trials resulted in "cycling."

- Fictitious play; each firm believes the next play of its competitors will be a random draw from the set of tuples of plays observed since the regime change (and best responds accordingly).

Note: here we consider forming beliefs about competitor's actions. An alternative would have been to consider "learning" about the outcomes of one's actions (that is I have a perception of the returns to different actions and update those beliefs). We return to this second formulation when we come to reinforcement learning.

Each run is stopped when we have converged to a single allocation, where convergence is defined as having remained in the same allocation state for 50 iterations. This location was viewed as a "rest point" of the process. Note that *all* rest points are Nash equilibria of the game where each agent knows its mean costs. Table 2 provides the fraction of rest points at various equilibria for the different cost specifications. We tried different mean cost-shocks and different coefficient variations for those shocks.

Note that

- The variance in the cost shocks can cause a *distribution* of rest points from a given initial condition.
- Apparently their is a dependence of the distribution of the equilibria on belief formulation process. This is troubling because of the lack of evidence on the empirical relevance on how one forms beliefs.
- On the brighter side, it appears that the distribution of the number of ATMs from the lower cost specifications always stochastically dominated those from the higher cost specifications.

7 A Very Brief Introduction to the Econometrics of Inequality Estimators.

Take $x_i \in \mathcal{J}_i$ and let $h : X \to \mathcal{R}_+$ be a positive function of x_i. Then if $\{d_i'\}$ are feasible choices when d_i is chosen what are theory gives us

$$E\left[\sum_i \Delta r(d_i, d_i', d_{-i}, z_i^0, \theta) \otimes h(x_i)\right] \geq 0, \quad at \;\; \theta = \theta_0.$$

We will assume the observations are grouped into markets and that the difference between expectations and realizations are not correlated across observations in different markets.

Estimator. Form sample analog and looks for values of θ that satisfy these *moment inequalities.*

Details Estimation.

Assume J markets $(j = 1, \ldots, J)$, with observations on (z, x, d) for all active agents. The markets' observations are independent draws from a population of markets with a distribution, say $\mathcal{P}$, that respects our two assumptions.

Sample Moments.
Let

$$m(z^j, d^j, x^j, \theta) =$$
$$\frac{1}{n_j} \sum_i \Delta r^j(d_i^j, d', d_{-i}^j, z_i^j, \theta) \otimes h(x_i^j)$$

where n_j is the number of agents in market j, and

$$m(P_J, \theta) = \frac{1}{J} \sum_{j=1}^{J} m(z^j, d^j, x^j, \theta)$$

while

$$\Sigma(P_J, \theta)$$

is the sample variance of these moments.

Population Moments.
Let

$$m(\mathcal{P}, \theta) = Em(\cdot, \theta),$$

and

$$\Sigma(\mathcal{P}, \theta) = Var(m(\cdot, \theta)).$$

Our assumptions imply

$$m(\mathcal{P}, \theta_0) \geq 0.$$

Estimator. If $f(\cdot)_- \equiv min(0, f(\cdot))$ then

$$\Theta_J = \arg\min_\theta \|m(P_J, \theta)_-\|.$$

Note that:

- Θ_J may be a set (rather than a point).
- Not be surprised if there is no value of θ that satisfies $m(P_J, \theta) \geq 0$ in finite samples. E.g.

$$m(\mathcal{P}, \theta_0) = 0 \ \Rightarrow$$

$$Pr\{m(P_J, \theta_0) \geq 0\} \rightarrow_{(M \rightarrow \infty)} 0.$$

Inference. Under standard conditions $\hat{\Theta}_J$ converges to Θ_0 (for standard norms defined on sets). There are at least three ways to obtain some indication of the precision of your set valued estimator. One can

- Find a set of points which covers the true θ_0 with required probability
- Find a set of points which covers the true identified set (Θ_0) with required probability
- Define $\underline{\theta}_{1,0} = inf_{\theta_1}\{\theta : \theta \in \Theta_0\}$ and $\overline{\theta}_{1,0}$ analogously. Find an interval which covers($\underline{\theta}_{1,0}, \overline{\theta}_{1,0}$) with required probability.

I am going to focus on the third, but presumably if you wanted to do either of the first two you would project them down onto the various axis when reporting results. The original paper on the econometrics of all this was by Chernozhukov, Hong and Tamer (2007, *Econometrica*) and they developed the first two techniques. For more on all of this you can look at the series of papers by Andrews and his co-authors (most of which are in, or are forthcoming, in *Econometrica*)[7]. Use of the third technique dates to PPHI.

We look for limiting distributions for the sample analogues of the lower and upper bounds of each component of θ_0, say $[\hat{\underline{\theta}}_{0,1}, \hat{\overline{\theta}}_{0,1}]$, and we will assume that there are unique vectors $(\underline{\theta}, \overline{\theta})$ associated with $(\underline{\theta}_{1,0}, \overline{\theta}_{1,0})$. We obtain those bounds from the limiting distribution of each of $\hat{\underline{\theta}}_{0,1}$ and $\hat{\overline{\theta}}_{0,1}$ separately, using the following inequality

$$\Pr\left\{\theta_{0,1} \in [\hat{\underline{\theta}}_{0,1}, \hat{\overline{\theta}}_{0,1}]\right\} \ \geq \ \Pr\left\{[\underline{\theta}_{0,1}, \overline{\theta}_{0,1}] \subset [\hat{\underline{\theta}}_{0,1}, \hat{\overline{\theta}}_{0,1}]\right\}$$
$$\geq \ 1 \ - \ \Pr\left\{\hat{\underline{\theta}}_{0,1} > \underline{\theta}_{0,1}\right\} \ - \ \Pr\left\{\hat{\overline{\theta}}_{0,1} < \overline{\theta}_{0,1}\right\}.$$

Note that a choice of $\hat{\underline{\theta}}_{0,1}$ and $\hat{\overline{\theta}}_{0,1}$ that sets the far right expression to $1 - \alpha$ is clearly conservative for $1 - \alpha$ level coverage for both $\theta_{0,1}$, and for the interval $[\underline{\theta}_{0,1}, \overline{\theta}_{0,1}]$.

Inference: Linear Case. The linear case

$$m(P_J, \theta) = Z_J \, \theta - W_J.$$

Then Θ_J and the "identified set",

$$\Theta_0 = \{\theta : \mathcal{Z} \, \theta \geq \mathcal{W}, \ \theta \in \Theta\}$$

[7] An accessible version of how to actually obtain the various estimates of precision can be found in my notes for when I teach econometrics, and I will make that available on request

are both convex and hence easy to analyze. Focus on the problem of finding confidence intervals for the k^{th} components of θ, θ_k. Let

$$\Theta_{k,0} = \{\theta_k : \theta \in \Theta_0\} = [\underline{\theta}_{k,0}, \overline{\theta}_{k,0}],$$

and

$$\Theta_{k,J} = \{\theta_k : \theta \in \Theta_J\} = [\underline{\theta}_{k,J}, \overline{\theta}_{k,J}].$$

Note that $\Theta_{k,J}$ easy to compute (fmincom from matlab is sometimes used, but experience indicates that a simplex method will do better).

To obtain distribution of estimators

- Simulate draws from a normal centered at (Z_J, W_J) with covariance matrix equal to the sample covariance of $\{(Z_{j,J}, W_{j,J})\}_j$ times $\sqrt{2lnlnJ}$.
- If the draw generates a set estimator, evaluate the bounds functions $\underline{\theta}_k$ and $\overline{\theta}_k$ at the values of the draws.
- If the draw generates a point, drop moments, in order of $(Z_{ns}\theta - W_{ns})_+$ until one gets a set, and then record the bounds.
- Repeat this procedure, obtain a distribution for the bounds.

Note. This is an active area of econometric research; see also references cited at the beginning of these notes.

8 Another example: Holmes 2011.

- Q: What are the benefits of store density for Walmart?
- Walmart diffusion: https://www.youtube.com/watch?v=EGzHBtoVvpc
- Idea: Walmart is vertically integrated into distribution (it owns both retail and regional distribution and grocery distribution centers). Close stores allows for savings on distribution/trucking costs; however, cost is cannibalization.
- Dynamic structural model of Walmart store roll out (1962–2005). Perturbation approach: Walmart could have changed what it did at each point in time; model assumes that observed choices were (in expectation) optimal.
- Abstracts from oligopolistic interactions (**?**) and instead focuses on dynamics and cannibalization.
- Focus of paper: Can ask how many and where to put new Walmarts/supercenters, and how many and where to put new distribution centers. Focuses on where to put new Walmarts/supercenters conditioning on the other decisions.

TABLE I

SUMMARY STATISTICS OF STORE-LEVEL DATA[a]

Store Type	Variable	N	Mean	Std. Dev.	Min	Max
All	Sales ($millions/year)	3,176	70.5	30.0	9.1	166.4
Regular Wal-Mart	Sales ($millions/year)	1,196	47.0	20.0	9.1	133.9
Supercenter	Sales ($millions/year)	1,980	84.7	25.9	20.8	166.4
All	Employment	3,176	254.9	127.3	31.0	812.0
Regular Wal-Mart	Employment	1,196	123.5	40.1	57.0	410.0
Supercenter	Employment	1,980	333.8	91.5	31.0	812.0

[a]End of 2005, excludes Alaska and Hawaii. Source: Trade Dimensions retail data base.

8.1 Data

1. ACNielsen estimates of store-level sales for all stores open at the end of 2005.
2. Opening dates of four types of Walmart facilities.
3. Demographic information from 1980, 1990, and 2000 US Census at *block group*, finer than Census tract. Population density from circle of 5mi radius centered at block group.
4. Local wages and rents from US Census Bureau Cuonty Business Patterns.
5. Walmart annual reports which include information on annual sales for earlier years (helpful to understand cannibalization effect).

8.2 Model

- Stores can stock food f or general merchandise g. Stores are either regular (only g) or supercenters (both f and g).
- Locations indexed by $l = 1, \ldots, L$, where $d_{l,l'}$ is distance between locations. $\mathcal{B}_t^W$ is set of locations with a Walmart, $\mathcal{B}_t^S \subseteq \mathcal{B}_t^W$ are supercenter sets.
- $R_{jt}^g(\mathcal{B}_t^W)$ are general merchandise sales revenues of store j at time t; $R_{jt}^f(\mathcal{B}_t^W)$ are food sales of j (if j is a supercenter). Fixed gross margin μ so general merchandise profits for store j is $\mu R_{jt}^g(\mathcal{B}_t^W)$.
- Costs:
  1. $DistributionCost_{jt} = \tau d_{jt}^g + \tau d_{jt}^f$ where d is distance to closest general or food distribution center (GDC or FDC); τ is cost per mile per period of servicing the store. Assumed to be the same for food and general.
  2. Variable Costs: fixed proportion of revenue, $\nu^{Labor}, \nu^{Land}, \nu^{Other}$
  3. Fixed Costs: $c(Popden_j) = \omega_0 + \omega_1 \ln(Popden_j) + \omega_2 \ln(Popden_j)^2$ depends on location population density (urban locations disadvantageous).
- Assumptions: Fixed discount factor $\beta = .95$; exogeneous productivity growth ρ_t. Walmarts don't close once opened.

$$\mathcal{B}_t^W = \mathcal{B}_{t-1}^W + \mathcal{A}_t^W$$
$$\mathcal{B}_t^S = \mathcal{B}_{t-1}^S + \mathcal{A}_t^S$$

where $\mathcal{A}_t$ are new stores of a given type opened at period t.

- Let $a = (\mathcal{A}_1^W, \mathcal{A}_1^S, \mathcal{A}_2^W, \mathcal{A}_2^S, \ldots)$. Holding fixed number of new stores, will focus on optimal location decision and hence actions that coincide with the number of stores actually opened.

$$\max_a \sum_{t=1}^{\infty} (\rho_t \beta)^{t-1} \Big[\sum_{j \in \mathcal{B}^W} [\pi_{jt}^g - c_{jt}^g - \tau d_{jt}^g] + \sum_{j \in \mathcal{B}_t^S} [\pi_{jt}^f - c_{jt}^f - \tau d_{jt}^f] \Big] \tag{6}$$

where $\pi_{jt}^e = \mu R_{jt}^e - Wage_{jt} Labor_{jt}^e - Rent_{jt} Land_{jt}^e - Other_{jt}^e$ for $e \in \{f, g\}$.

NB: sunk costs do not affect the objective of where to locate stores *conditional* on number of stores that are opened.

8.2.1 Operating Profits

Consumer demand:

- Consumers can choose between outside option and any Walmart within 25mi radius ($\bar{\mathcal{B}}_l^W$).

$$u_0 = b(Popden_l) + LocationChar_l \alpha + \varepsilon_0$$
$$b(Popden_l) = \alpha_0 + \alpha_1 \ln(\underline{Popden}) + \alpha_2 (\ln(\underline{Popden}))^2$$

where $\underline{Popden} = \max\{1, Popden\}$, $Popden$ is thousands of people within 5mi radius, and $b'(\cdot) \geq 0$ to capture idea that denser markets have better outside options.

Utility of Walmart given by:

$$u_{lj} = -[\xi_0 + \xi_1 \ln(\underline{Popden})] Distance_{lj} + StoreChar_j \gamma + \varepsilon_j$$

- Operating revenues: $R_j^g = \sum_{l: j \in \bar{\mathcal{B}}_l^W} \lambda^g \times p_{jl}^g \times n_l$ where λ^g is spending per consumer and n_l are consumers at location l. Food defined similarly.

- Measurment error $\eta_j^{Sales} = \ln(R_j^{Data}) - \ln(R_j^g(\Psi))$ where Ψ is parameter vector from demand model; assume that measurement error is iid normal and estimate using MLE.

- Add in data from annual reports: "As we continue to add new stores domestically, we do so with an understanding that additional stores take sales away from existing units. We estimate that comparative store sales in FY 2004, 2003, 2002 were negatively impacted by the opening of new stores by approximately 1%."

- Can calculate "cannibalization rate" for a given store by computing demand if no stores were opened that year, and after the observed openings occurred. Seems to match; however, insofar upper bound on degree of density economies closely connected to degree of cannibalization, also explore results when cannibalization is constrained in 2005 to be 1. To be conservative, use this result.

TABLE V

CANNIBALIZATION RATES, FROM ANNUAL REPORTS AND IN MODEL[a]

Year	From Annual Reports	Demand Model (Unconstrained)	Demand Model (Constrained)
1998	n.a.	.62	.48
1999	n.a.	.87	.67
2000	n.a.	.55	.40
2001	1	.67	.53
2002	1	1.28	1.02
2003	1	1.38	1.10
2004	1	1.43	1.14
2005	1	1.27	1.00[b]

[a]Source: Estimates from the model and Wal-Mart Stores, Inc. (1971–2006) (Annual Reports 2004, 2006).

[b]Cannibalization rate is imposed to equal 1.00 in 2005.

[12]Wal-Mart's fiscal year ends January 31, so the fiscal year corresponds (approximately) to the previous calendar year. For example, the 2006 fiscal year began February 1, 2005. In this paper, I aggregate years like Wal-Mart (February through January), but I use 2005 to refer to the year beginning February 2005.

8.3 Bounding Density Economies

- Need to recover $\theta = (\tau, \omega_1, \omega_2)$; τ is Walmart economies of density, ω are diseconomies of population density.
- Linear Moment Inequality Framework: let there be M linear inequalities where:

$$y_a \geq x_a' \theta, \qquad a \in \{1, 2, \ldots, M\}$$

is assumed to hold for θ^0. With K non-negative instruments for each a:

$$z_{ak} y_a \geq z_{ak} x_a' \theta \; \forall \; a, k \tag{7}$$

Assume measurement error (ν_1) on y_a so we observe $\tilde{y}_a = y_a + \eta_a$ where $E[\eta_a | x_a, z_{ak}] = 0$. Let $m_k(\theta) = E[z_{ak}\tilde{y}_a] - E[z_{ak}x_a']\theta$; then $m_k(\theta) \geq 0$ at true parameter vector. Let $Q(\theta) = \sum_{k=1}^{K} (\min\{0, m_k(\theta)\})^2$.

Sample analogues:

$$\tilde{m}_k(\theta) = \sum_{a=1}^{M} \frac{z_{ak}\tilde{y}_a}{M} - \sum_{a=1}^{M} \frac{z_{ak}x_a'}{M}\theta$$

$$\tilde{Q}(\theta) = \sum_{k=1}^{K} (min\{0, \tilde{m}_k(\theta)\})^2$$

and let identified set $\hat{\Theta}^I = \arg\min_{theta} \tilde{Q}(\theta)$. (Need to average across M inequalities to remove measurement error).

TABLE IV

PARAMETER ESTIMATES FOR DEMAND MODEL

Parameter	Definition	Unconstrained	Constrained (Fits Reported Cannibalization)
λ^g	General merchandise spending per person (annual in \$1,000)	1.686 (.056)	1.938 (.043)
λ^f	Food spending per person (annual in \$1,000)	1.649 (.061)	1.912 (.050)
ξ_0	Distance disutility (constant term)	.642 (.036)	.703 (.039)
ξ_1	Distance disutility (coefficient on $\ln(\underline{\text{Popden}})$)	−.046 (.007)	−.056 (.008)
α	Outside alternative valuation parameters		
	Constant	−8.271 (.508)	−7.834 (.530)
	$\ln(\underline{\text{Popden}})$	1.968 (.138)	1.861 (.144)
	$\ln(\underline{\text{Popden}})^2$	−.070 (.012)	−.059 (.013)
	Per capita income	.015 (.003)	.013 (.003)
	Share of block group black	.341 (.082)	.297 (.076)
	Share of block group young	1.105 (.464)	1.132 (.440)
	Share of block group old	.563 (.380)	.465 (.359)
γ	Store-specific parameters		
	Store age 2 + dummy	.183 (.024)	.207 (.023)
σ^2	Measurement error	.065 (.002)	.065 (.002)
N		3,176	3,176
Sum of squared error		205.117	206.845
R^2		.755	.753
(Likelihood)		−155.749	−169.072

- Mapping this to Walmart:
  - Walmart objective in (6) is linear in θ. Define $y_a = \Pi(a^o) - \Pi(a)$ for:

$$\Pi(a) = \sum_{t=1}^{\infty} (\rho_t \beta)^{t-1} \Big(\sum_{j \in \mathcal{B}_t^W(a)} \pi_{jt}^g(a) + \sum_{j \in \mathcal{B}^S(a)} \pi_{jt}^f(a) \Big) \tag{8}$$

    y_a represents incremental PDV of operating profit from implementing observed policy a^o instead of deviation a. x_a will include the PDV of differences in distribution distances, $\ln(Popden)$, and $\ln(Popden)^2$ from policy a^o versus a.
  - Dealing with measurement error and demand estimation error: see paper.

TABLE VII

DISTRIBUTION OF VARIABLE INPUT COSTS[a]

Estimated 2005 Labor Costs			
Quartile	Store Location	Annual Payroll per Worker ($)	Wages as Percentage of Sales
Minimum	Pineville, MO	12,400	4.5
25	Litchfield, IL	19,300	7.0
50	Belleville, IL	21,000	7.6
75	Miami, FL	23,000	8.3
Maximum	San Jose, CA	37,900	13.7
Estimated 2005 Land Value–Sales Ratios			
Quartile	Store Location	Index of Residential Property Value per Acre ($)	Land Value as Percentage of Sales
Minimum	Lincoln, ME	1,100	.0
25	Campbellsville, KY	32,100	1.2
50	Cleburne, TX	67,100	2.4
75	Albany, NY	137,300	5.0
Maximum	Mountain View, CA	1,800,000	65.0

[a]Percentiles of distribution are weighted by sales revenue.

– Choice of deviations: Focus on *pariwise resequencing* – i.e., deviations where opening dates of pairs of stores are reordered. E.g., store 1 in 1962 and store 2 in 1964; instead open store 1 in 1964 and store 2 in 1962.

8.4 Results

1. Making a store closer to a distribution center by 1 mile yields a benefit of $3,500 / year. Savings in trucking costs estimated to be about 850/year, so total savings 4 times larger. Difference includes "value Walmart places on the ability of responding quickly to demand shocks."

2. If all 5K Walmart stores were each 100 miles further from their distribution centers, Walmarts costs would increase by almost $2B per year.

TABLE IX

DEFINITIONS OF DEVIATION GROUPS[a]

Deviation Category	Deviation Group	Description (Store j' Flips with Store j)
Store density decreasing		Find the set of stores, $S = \{j, t_j \geq t_j^{\text{state}} + 10\}$. For each $j \in S$, find all j', where (i) $t_{j'} \geq t_j + 3$, (ii) j' is in a different state than j, and (iii) $t_{j'} \leq t_{j'}^{\text{state}} + 4$. Take all of these and further classify by group on the basis of ΔD_a as follows:
	1	$-.75 \leq \Delta D_a < 0$
	2	$-1.50 \leq \Delta D_a < -.75$
	3	$\Delta D_a < -1.50$
Store density increasing		Find the set of stores, $S = \{j, t_j \leq t_j^{\text{state}} + 5\}$. For each $j \in S$, find all j', where (i) $t_{j'} \geq t_j + 3$, (ii) j' is in a different state than j, and (iii) $t_{j'} \geq t_{j'}^{\text{state}} + 10$. Take all of these and further classify by group on the basis of ΔD_a as follows:
	4	$0 < \Delta D_a \leq .75$
	5	$.75 < \Delta D_a \leq 1.50$
	6	$1.50 < \Delta D_a$
Population density changing		Take pairs of stores (j, j') opening in the same state, where $t_j \leq t_{j'} + 2$. Classify based on $\underline{\text{Popden}}_j$ (in units of 1,000 people within 5-mile radius). Define density classes 1, 2, 3, and 4 by $\underline{\text{Popden}}_j < 15$, $15 \leq \underline{\text{Popden}}_j < 40$, $40 \leq \underline{\text{Popden}}_j < 100$, and $100 \leq \underline{\text{Popden}}_j$.
	7	j in class 4, j' in class 3
	8	j in class 3, j' in class 2
	9	j in class 2, j' in class 1
	10	j in class 1, j' in class 2
	11	j in class 2, j' in class 3
	12	j in class 3, j' in class 4

[a]Notes: The table uses the following notation: t_j is the opening date of store j, t_j^{state} is the opening date of the first store in the state where j is located, ΔD_a is the present value of the increment in distribution distance miles (in 1,000s) from doing the actual policy $a°$ instead of deviating and doing a. In words, to construct group 1, take the set of all stores opening when there is at least one store in their state that is 10 years old or more. For each such store, find alternative stores that open 3 or more years later in different states, where Wal-Mart has been in the different state no more than 4 years when the alternative store opens. Openings for general merchandise stores and food stores are considered two different opening events. In cases where a supercenter opens from scratch rather than as a conversion of an existing Wal-Mart, there are two opening events. In all the pairs considered, a general merchandise opening is paired with another general merchandise opening, and a food opening with another food opening.

TABLE X

SUMMARY STATISTICS OF DEVIATIONS BY DEVIATION GROUP

Deviation Group	Brief Description of Group	Number of Deviations	Mean Values: $\Delta\tilde{\Pi}$ (Millions of 2005 Dollars)	ΔD (Thousands of Miles)	ΔC_1 (log Popden)	ΔC_2 (log Popden^2)
	Store density decreasing					
1	$-.75 \leq \Delta D < 0$	64,920	−2.7	−.4	−.6	−3.0
2	$-1.50 \leq \Delta D < -.75$	61,898	−3.6	−1.1	−1.5	−9.0
3	$\Delta D < -1.50$	114,588	−4.7	−3.0	−3.4	−22.2
	Store density increasing					
4	$0 < \Delta D \leq .75$	158,208	3.0	.3	−1.9	−17.2
5	$.75 < \Delta D \leq 1.50$	34,153	3.7	1.0	−3.6	−28.9
6	$1.50 < \Delta D$	16,180	5.9	2.1	−4.8	−37.7
	Population density changing					
7	Class 4 to class 3	7,048	1.2	.0	3.2	31.1
8	Class 3 to class 2	10,435	3.7	.0	3.4	25.7
9	Class 2 to class 1	14,399	5.3	−.1	3.5	19.3
10	Class 1 to class 2	12,053	−2.4	.0	−3.4	−19.3
11	Class 2 to class 3	14,208	.6	−.1	−3.9	−29.4
12	Class 3 to class 4	14,877	2.5	.0	−4.6	−44.9
All	Weighted mean	522,967	−.2	−.6	−2.1	−15.6

TABLE XI

BASELINE ESTIMATED BOUNDS ON DISTRIBUTION COST τ[a]

	Specification 1 Basic Moments (12 Inequalities)		Specification 2 Basic and Level 1 (84 Inequalities)		Specification 3 Basic and Levels 1, 2 (336 Inequalities)	
	Lower	Upper	Lower	Upper	Lower	Upper
Point estimate	3.33	4.92	3.41	4.35	3.50	3.67
Confidence thresholds						
With stage 1 error correction						
PPHI inner (95%)	2.69	6.37	2.89	5.40	3.01	4.72
PPHI outer (95%)	2.69	6.41	2.86	5.45	2.97	5.04
No stage 1 correction						
PPHI inner (95%)	2.84	5.74	2.94	5.11	3.00	4.62
PPHI outer (95%)	2.84	5.77	2.93	5.13	2.99	4.97

[a] Units are in thousands of 2005 dollars per mile year; number of deviations $M = 522{,}967$; number of store locations $N = 3{,}176$.