

Lecture 4: NBERMetrics.

Ariel Pakes

July 17, 2012

On The Use of Moment Inequalities in Discrete Choice Models.

From the point of view of consumer theory the work on estimating preference parameters from moment inequalities is a follow up to theory literature on “*revealed preference*” begun by Paul Samuelson (1938, *Economica*); see the review by Hal Varian (2005, in *Samuelsonian Economics and the 21st Century*). The difference between the theory treatment of revealed preference, and the approach I will sketch here is our interest in estimation and therefore our attention to the sources of disturbances that enter the revealed preference approach when it is brought to data. The approach used here dates to Pakes, Porter, Ho and Ishii (2011, *working paper*) henceforth PPHI, and Pakes (2011, *Econometrica*).

It differs from standard empirical models of choice by working directly with the inequalities which define optimal behavior. I.e. we see how far we can get by just assuming that we know a

counterfactual which agents considered and discarded, and then assuming that, at least on average, the utility from the actual choice should be larger than the utility from the counterfactual. Only certain values of the parameter vector will make this statement true, and we accept any parameter values which insure that the average difference is positive.

Typically if one value satisfies the inequalities, so will values close to it. So there will be many values of the parameter vector that satisfy the inequalities and we will obtain a “set” of acceptable values. If the model is correct, what we will know is that the true value is in the set (at least asymptotically).

I have found it easiest to explain what is going on in this approach by starting out a simple single agent example.

Single Agent Example: Due to M. Katz (2007) Estimate the costs shoppers assign to driving to a supermarket. This is a topic of importance to the analysis of zoning regulations, public transportation projects, and the like. Moreover, it has proven difficult to analyze empirically with standard choice models because of the complexity of the choice set facing consumers: all possible bundles of goods at all possible supermarkets.

We shall see that though large choice sets make discrete choice analysis difficult, the larger the choice set the easier it is to use the moment inequality approach.

Assume that the agents’ utility functions are additively separable functions of;

- utility from the basket of goods bought,

- expenditure on that basket, and
- drive time to the supermarket.

Let $d_i = (b_i, s_i)$ the vector of decisions made by the agent, $b_i = b(d_i)$ be the basket of goods bought, $s_i = s(d_i)$ is the store chosen, and z_i are individual characteristics

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

where $W(\cdot)$ is the utility from the bundle of goods bought, $e(\cdot)$ provides expenditure, $dt(\cdot)$ provides drive time, and I have used the free normalization that comes from the equivalence of affine transforms on expenditure (so the cost of drive time are in dollars).

Standard Discrete Choice Model's For This Problem. There are really two aspects that make a behavioral (or structural) model of this decision difficult, and two approaches to the analysis.

A standard full information structural model would assume that the agent knows the prices of all bundles of goods at all stores, and makes its choice by

- computing the basket that maximizes the utility at each store, and substituting that into the $U(\cdot)$, and then
- maximizing over stores.

This has both behavioral and computational aspects which are troubling.

- Behaviorally it endows the agent with both; an incredible amount of information, and massive computational abilities.
- Computationally it requires specification of and computation of the value of a huge choice set, and if the $U(b_i, \theta)$ is specified in a reasonably rich way, simultaneous estimation of a large number of parameters. Both of these make obtaining consistent estimates a difficult task.

Probably a more realistic way of looking at the problem would be through the lenses of a two period model.

- In the first period the agent choses which store to travel to, perhaps without full knowledge of the prices at each store.
- In the second period the agent, having gone to the store chosen, choses the bundle of goods it wishes to buy from the goods available at the store.

This does not ameliorate the computational complexity of the problem (see below). However it does get rid of the behavioral problem. On the other hand, it replaces the behavioral problem with a specification problem. That is to proceed in this way the econometrician

- needs to specify the agent's prior probability for each possible price at each store, and
- then compute the integral of the bundle chosen given each possible price vector at each store

Econometricians never see priors, and seldom have knowledge of what the agent can condition on when it formulates its prior.

So the allowance for incomplete information replaces the behavioral problem with a specification problem. Moreover if anything, it adds to the computational complexity of the analysis by requiring the computation of expectations¹.

Behavioral Model vs. Descriptive Summary Statistics. I have focused here on estimating the parameters of the behavioral model. I did this because the interest was in a parameter vector which was then to be used for policy analysis (which presumably involves counterfactuals). If we were not interested in the behavioral parameters, but rather were interested in summarizing the data on store choice in an intuitive way (say a way that subsequent research might use to build stylized models for the choice) , there is a form of a standard discrete choice model that makes sense (at least for the single agent problems we are concerned with today). We would

- group choices of baskets of goods somehow,
- project the utility from the basket down on “variables of interest” which leaves a residual which is orthogonal to those variables.
- evaluate the possible baskets at different stores.

Then if we were willing to assume the distribution of the residual satisfied some parametric form we would be back to the familiar discrete choice model. However

- the coefficients could not be relied upon to provide an adequate approximation to what would happen were we to

¹One way of getting around these problems is to obtain survey measures of expectations (see the Manski, 2001, review in *Econometrica*), but this is often impracticable.

change a variable of interest (we come back to an example of this below), and

- it is not clear what one is doing in using a more complicated discrete choice model, for example a random coefficients model, in this context.

The Revealed Preference Approach. We compare the utility from the choice the individual made to that of an alternative feasible choice. Our theoretical assumption is that the agent expects this difference to be positive.

The analogue to the “sample design” question here is:

Which alternative should the econometrician choose?

The answer depends on the parameter of interest. Given this interest, the larger the choice set, the easier it will be to find an alternative that does this. This is the sense in which increasing the size of the choice set can not but help the moment inequality analysis.

Our interest is in analyzing the cost of drive time, so we chose an alternative which will allow us to analyze that without making the assumptions required to estimate $W(b, z, \theta)$.

For a particular d_i , choose $d'(d_i, z_i)$ to be the purchase of

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

Notice that choosing the same basket at the alternative store is dominated by choosing the optimal basket for that store which, at least before going to the store, is revealed to be inferior to choosing the optimal basket for the store chosen. So transitivity of preferences gives us the desired inequality.

With this alternative we 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. Notice also that

- We need not specify the whole choice set, another issue which is typically a difficult specification issue,
- The alternative chosen differs with the agent's choice and characteristics.

The Analysis. Let $\mathcal{E}(\cdot)$ be the *agent's* expectation operator, and for any function $f(x, d)$, let

$$\Delta f(x, d, d') \equiv f(x, d) - f(x, d').$$

Then the behavioral assumption we will make is that if d_i was chosen and $d'(d_i, z_i)$ was feasible when that decision was made

$$\begin{aligned} & \mathcal{E}[\Delta U(d_i, d'(d_i, z_i), z, \theta) | \mathcal{J}_i] = \\ & -\mathcal{E}[\Delta e(d_i, d'(d_i, z_i)) | \mathcal{J}_i] - \theta_i \mathcal{E}[\Delta dt(d_i, d'(d_i, z_i)) | \mathcal{J}_i] \geq 0. \end{aligned}$$

Notice that

- I have not assumed that the agent's perceptions of prices are "correct" in any sense; I will come back to what I need here below
- I have not had to specify the agent's priors or the information those priors rely on; and so have avoided the difficult specification issue referred to above.

The Inequalities.

We develop inequalities to be used in estimation for two separate cases.

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 (in terms of our previous notation, there is no random coefficient on drive time). Assume that

$$N^{-1} \sum_i \mathcal{E}[\Delta e(d_i, d'(d_i, z_i))] - N^{-1} \sum_i \Delta e(d_i, d'(d_i, z_i)) \rightarrow_P 0,$$

$$N^{-1} \sum_i \mathcal{E}[\Delta dt(d_i, d'(d_i, z_i))] - N^{-1} \sum_i \Delta dt(d_i, d'(d_i, z_i)) \rightarrow_P 0$$

where \rightarrow_P denotes convergence in probability. This would be true if, for example, agents were correct on average. Recall that

our assumption on the properties of the choice implied that

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

so the assumption made above that agents do not err on average gives us

$$-\frac{\sum_i \Delta e(d_i, d'(d_i, z_i))}{\sum_i \Delta dt(d_i, d'(d_i, z_i))} \rightarrow_p \underline{\theta} \leq \theta_0.$$

If we would have also taken an alternative store which was closer to the individual (say alternative d'') then

$$-\frac{\sum_i \Delta e(d_i, d''(d_i, z_i))}{\sum_i \Delta dt(d_i, d''(d_i, z_i))} \rightarrow_p \bar{\theta} \geq \theta_0.$$

and we would have consistent estimates of $[\underline{\theta}, \bar{\theta}]$ which bound θ_0 .

Case 2: $\theta_i = (\theta_0 + \nu_i)$, $\sum \nu_i = 0$. This case allows for a determinant of the cost of drive times (ν_i) that is known to the agent at the time the agent makes its decision (since the agent conditions on it when it makes its decision), but is not known to the econometrician. It corresponds to the breakdown used in the prior lecture of a coefficient which is a function of observables and an unobservable term (or the random coefficient); i.e.

$$\theta_i = z_i \beta_z + \nu_i.$$

Of course were we to actually introduce additional observable determinants of θ_i we would typically want to introduce also additional inequalities. We come back to this below, but for simplicity we now concentrate on estimating an (unconditional)

mean; i.e. θ_0 . Though this specification looks the same as that in the last lecture, the assumptions on the random terms made here and the conclusions we can draw from our estimator will be different now.

Now provided $dt(d_i)$ and $dt(d'(d_i, z_i))$ are known to the agent when it makes its decision

$$-\mathcal{E}[\Delta e(d_i, d'(d_i, z_i))] - (\theta_0 + \nu_i)[\Delta dt(d_i, d'(d_i, z_i))] \geq 0$$

which, since $\Delta dt(d_i, d'(d_i, z_i)) \leq 0$, implies

$$\mathcal{E}\left[\frac{-\Delta e(d_i, d'(d_i, z_i))}{\Delta dt(d_i, d'(d_i, z_i))} - (\theta_0 + \nu_i)\right] \leq 0.$$

So provided agents' expectations on expenditures are not “systematically” biased

$$\frac{1}{N} \sum_i \left(\frac{\Delta e(d_i, d'(d_i, z_i))}{\Delta dt(d_i, d'(d_i, z_i))} \right) \rightarrow_P \underline{\theta} \leq \theta_0.$$

and we could get a lower bound as above.

There are two points about this latter derivation that should be kept in mind.

- First we did not need to assume that ν_i is independent of z_i . This would have also been true in a richer model where we assumed $\theta_i = z_i \beta_z + \nu_i$. This is particularly useful in models for purchase of a good that has a subsequent cost of use like a car or an air conditioner. Then utilization (or expected utilization) would be a right hand side variable and one would think that unobservable determinants of the

importance of efficiency in use in product choice would be correlated with the determinants of utilization (in the car example, consumers who care more about miles per gallon are likely to be consumers who expect to drive further).

- Relatedly we did not have to assume a distribution for the ν_i . The flip side of this is that neither have we developed an estimator for the variance of the random term. Though, with additional assumptions we could develop such an estimator, in this lecture I am not going to do that (though variance in the observable determinants of the aversion will be considered below).

Case 1 vs. Case 2.

- Case 1 estimates using the ratio of averages, while case 2 estimates by using the average of the ratios. Both of these are trivially easy to compute.
- Case 2 allows for unobserved heterogeneity in the coefficient of interest and does not need to specify what the distribution of that unobservable is. Case 1 ignores the possibility of unobserved heterogeneity in tastes.
- 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 to be negatively correlated with price itself.)
- it assumes that the agents knew the goods available in the store and their prices exactly when they decided which store to choose (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....

For these reasons, as well as the aggregation, its estimates are the estimates of a model which generates “summary statistics”

models as discussed above (colloquially referred to as a reduced form model), and should not be interpreted as causal.

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 per hour 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. The inequality estimators were obtained from differences between the chosen store and four different counterfactual store choices (chosen to reflect price and distance differences with the chosen store). Each comparison was interacted with positive functions of twenty six “instruments” , producing over a hundred moment inequalities. We come back to a more formal treatment of instruments below but for now I simply note that what we require of each instrument, say $h(x)$, is that it be

- known to the agent when it made its decision (so when the agent makes its decision the agent can condition on the instrument’s value)
- uncorrelated with the sources of the unobservable that the agent knew when it made its decision (in our case ν_i)
- and non-negative.

If these conditions are satisfied and

$$\mathcal{E}\left[\frac{-\Delta e(d_i, d'(d_i, z_i))}{\Delta dt(d_i, d'(d_i, z_i))} - (\theta_0 + \nu_i)\right] \leq 0.$$

then so will be

$$\mathcal{E}h(x_i)\left[\frac{-\Delta e(d_i, d'(d_i, z_i))}{\Delta dt(d_i, d'(d_i, z_i))} - (\theta_0 + \nu_i)\right] \leq 0.$$

So, again, provided agents' expectations are not “systematically” biased

$$\frac{1}{N} \sum_i \left(\frac{h(x_i) \Delta e(d_i, d'(d_i, z_i))}{\Delta dt(d_i, d'(d_i, z_i))} \right) \rightarrow_P \underline{\theta} \leq \theta_0.$$

and we have generated an additional moment inequality.

As is not unusual for problems with many more inequalities than bounds to estimate, the inequality estimation routine generated point (rather than interval) estimates for the coefficients of interest (there was no value of the parameter vector that satisfied all of the moment inequalities). However tests, that we come back to below, indicated that one could accept the null that this result was due to sampling error. The numbers beside the estimates are confidence intervals with the same interpretation as confidence intervals as we are used to; under the null the confidence interval will capture the true parameter value in 95% of samples. The confidence intervals presented here are from a very conservative procedure and could be tightened.

- Inequality estimates with

$$\theta_i = \theta_0 : .204 \text{ } [.126, .255]. \Rightarrow \$4/hour,$$

- Inequality estimates with

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

and other coefficients straighten out.

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

Katz only estimated the mean of the drive time coefficient conditional on observable demographics, and then used it, plus the distribution of demographics to evaluate policy counterfactuals. If one were after more details on the $W(\cdot)$ function, one would consider different counterfactuals (and presumably additional functional form restrictions). If one observed the same household over time and was willing to assume stability of preference parameters over time, many other options are available, some mimicing what we did with second choice data in the last lecture.

Estimation, Testing, and C.I.'s: An Introduction.

The finding that there is no value of the parameter vector that satisfies all the inequalities is not unusual in moment inequality problems with many inequalities. Consider the one parameter case.

When there are many moment inequalities there are many upper and lower bounds for that parameter. The estimation routine forms an interval estimate from the least upper and the greatest lower bound. In finite samples the moments distribute normally about their true value and as a result there will be a

negative bias in choosing the smallest of the upper bounds due to finite sample sampling error. Similarly there will be a positive bias in choosing the largest of the lower bounds. It is easy for these two to cross, even if the model is correctly specified. So there is an interest in building a test which distinguishes between the possibility that we obtain a point because of sampling error, and the possibility that we obtain a point because the model is misspecified.

We begin, however, by setting up the estimation problem a bit more formally. For more on what follows see Chernozhukov, Hong and Tamer (Econometrica, 2007), Andrews and Soares (Econometrica, 2010), and the articles cited above.

Our model delivers the condition that

$$\begin{aligned} & E [\Delta U(d_i, d'(d_i, z_i), \mathbf{z}_i, \theta_0) \otimes h(x_i)] \\ & \equiv E[m(d_i, z_i, x_i, \theta)] \geq 0, \text{ at } \theta = \theta_0. \end{aligned}$$

where both $\Delta U(\cdot)$ and $h(\cdot)$, and hence $m(\cdot)$, may be vectors and \otimes is the Kronecker product operator.

Estimator. Essentially we form a sample analog which penalizes values of θ that does not satisfy these conditions, but accepts all those that do. More formally the sample moments are

$$m(\mathbf{P}_n, \theta) = \frac{1}{n} \sum_i m(d_i, z_i, x_i, \theta)$$

and their variance is

$$\Sigma(\mathbf{P}_n, \theta) = Var(m(z_i, d_i, x_i, \theta)).$$

The corresponding population moments are $(m(\mathcal{P}, \theta), \Sigma(\mathcal{P}, \theta))$ and our model assumes

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

The set of values of θ that satisfy the inequalities is denoted

$$\Theta_0 = \{ \theta : m(\mathcal{P}, \theta) \geq 0 \},$$

and called the identified set. This is all we can hope to estimate.

Estimator. For now I am going to assume that Σ_0 , the covariance matrix of the moments is the identity matrix. This will lead to an estimator which can be improved upon by dividing each moment by an estimate of its standard error, and repeating what I do here, but that makes the notation more difficult to follow.

Recall that our model does not distinguish between values of θ that make the moments positive. All the model says is that at θ_0 the moments are positive (at least in the limit). So all we want to do is to penalize values of θ that lead to negative values of the moments. More formally if $f(\cdot)_- \equiv \min(0, f(\cdot))$ and for any vector f , $\|f\|$ denotes the sum of squares of f , then our estimator is

$$\Theta_n = \arg \min_{\theta} \|m(\mathbf{P}_n, \theta)_-\|.$$

This could be a set, or, as in Katz's case, it could be a single parameter value. Several papers prove that Θ_n converges to Θ_0 in an appropriate norm.

Measures of Precision. There are several different ways of conceptualizing measures of the precisions of your (set) estimator. We could attempt to

- get a confidence set for the set; i.e. a set which would cover the identified set 95% of the time (starts with Chernozhukov Hong and Tamer, *Econometrica* 2007)
- Get a confidence set for the point θ_0 (starts with Imbens and Manski, *Econometrica*, 2004).
- get confidence interval for intervals defined for a particular direction in the parameter space; simplest case is directions defined by each component of the parameter space so we get CI's which cover the different components of the parameter vector 95% of the time (analogous to reporting standard errors component by component, see PPHI, 2011)

There are also several ways to obtain CI's for each of the different concepts. I am going to give a conceptually simple method for getting a confidence set for the point θ_0 . The procedure I will go over will also be simple computationally when the θ vector is low dimensional. For high dimensional θ vectors you will often have to go to PPHI. For precision improvements on the techniques described here see Andrews and Guggenberger, (*Econometric Theory*, 2009), and the literature cited there.

Confidence Sets for the Point, θ_0 .

The procedure here tests each “possible value” of θ to see if it can reasonably be assumed to satisfy the moment inequalities.

To do this we have to start out by defining a grid of possible values say $\{\Theta_p\} = \{\theta_{p,j}, j = 1, \dots, J_p\}$. It is because we have to do the following procedure at every one of the points in the grid that this procedure often becomes impractical when there is a large dimensional parameter vector.

To accept a value of θ_p the value of the objective function when evaluated at θ_p (or $\|m(\theta_p, \mathbf{P}_n)_-\|$) must be within sampling error of zero. The problem is that the distribution of this object is not analytic. However if we are willing to assume a value for the mean of $m(\theta_p, \mathbf{P}_n)$, then it is easy to simulate a good approximation to the distribution of $\|m(\theta_p, \mathbf{P}_n)_-\|$, as the central limit theorem insures that $m(\theta_p, \mathbf{P}_n)$ distributes normally about that mean with a variance that is well approximated by $Var(m(\theta_p, \mathbf{P}_n))$.

Assuming temporarily that the mean $m(\theta_p, \mathbf{P}_n) = 0$ we construct the distribution of $\|m(\theta_p, \mathbf{P}_n)_-\|$ under that assumption. Take random draws from a normal distribution with mean zero and variance-covariance $Var(m(\theta_p, \mathbf{P}_n))$. For each random draw, say $\tilde{m}(\theta_p)$, form the Euclidean norm of its negative part, that is compute $\|\tilde{m}(\theta_p, \mathbf{P}_n)_-\|$, and keep track of the distribution of these numbers. Let $c(\theta_p)$ be the $(1 - \alpha)$ quantile of this distribution. $c(\theta_p)$ is the critical value for our test. Note that it is greater than the critical value that would obtain from a similar procedure but an assumption that the expectation of $m(\theta_p, \mathbf{P}_n)$ is any positive number.

Now go back to the original sample and compute the actual value of $\|m(\theta_p, \mathbf{P}_n)\|$. If it is less than $c(\theta_p)$, θ_p is accepted into the estimate of the confidence set for θ_0 . If not it is rejected.

Because we have used a critical value that is larger than the critical value for any other mean assumption that is consistent with the model, the probability of rejecting the null when it is true is less than α no matter the true value of the mean of $m(\theta_p, \mathbf{P}_n)$. If there is no value of θ_p that is accepted, we reject the null that the model is correctly specified.

Moment Inequalities vs. Standard Discrete Choice.

The limits of the moment inequality methods are still being actively explored. Its advantages are that

- It typically requires weaker behavioral assumptions than the standard theory. In particular it need not specify information sets and prior probabilities when it is likely that the agent does not have perfect information on the properties of the choice set, and allows agents to make errors, provided they are not consistently wrong. That is, it has a way of controlling for expectational errors provided they are mean zero.
- Something I have not stressed here, but is important in many applications, is that the fact that moment inequalities only use averages implies that it also controls for zero mean measurement errors (see Ho and Pakes 2012, *working paper*).
- It does not require specification of the entire choice set; just one counterfactual choice that the agent knew was feasible when it made the choice that it did make.

The biggest disadvantage of moment inequalities is that

- It has limited ability to handle variables that the agent does know when it makes its choice, but the econometrician does not observe.

In the Katz example we allowed for two types of unobservables that the agent knew when it made its decision, but we did not observe.

- One was the class of factors that determined the utility from the bundle of goods bought. In our notation we did not need to specify $W(b_i, z_i, \theta)$, so any determinant of that quantity could be unobserved and not impede the estimation algorithm.
- In case 2 we also did not observe a determinant of the value of drive time, our ν_i .

We were able to account for the determinants of the utility from the bundle of goods bought because of:

- the assumed additivity of the utility function in that bundle, and
- the fact that we had a very large choice set, a choice set that allowed us to hold that bundle constant at the same time as changing the dimension of the choice whose impact we wanted to investigate.

We were able to account for the an unobserved determinant of the importance of drive time that the agent knew but the econometrician did not, (our θ_i) because

- we could find a counterfactual which was linear in the effect of θ_i for *every* individual in the sample.

We then estimated the average effect of these variables.

Had we only been able to find such a transform for a subset of the consumers in the sample, say those that satisfy a particular condition, we would have had to deal with a selected sample of θ_i , and we would not have been able to estimate the mean of interest (at least not without either further assumptions, or a wider bound than provided here).

A More General Model. A more general model would have the agent maximizing

$$\mathcal{E}[U(d_i, z_i, \theta)] = \mathcal{E}[\tilde{W}(b_i, z_i, \theta_b)] - \mathcal{E}[\tilde{e}(b_i, s_i)] - \theta_i dt(s_i, z_i) + \epsilon_{i,s}$$

where

- a tilde over a variable indicates that it is random when the decision of which store to go to is taken and
- $\epsilon_{i,s}$ is a disturbance which differs by store that is known to the agent as it makes its store choice but not to the econometrician.

We observe the actual

$$W(\cdot) \quad (\text{not } \mathcal{E}[\tilde{W}(b_i, z_i, \theta_b)]), \quad \text{and} \quad e(\cdot) \quad (\text{not } \mathcal{E}[\tilde{e}(b_i, s_i)]).$$

The difference can be caused by either zero mean expectational or measurement error.

- The discrete choice model *does not allow* for these differences between the observed right hand side variables and their expectations. It *does however allow* for a parametric form for the distribution of $\epsilon_{i,j}$.
- The inequality model presented here *does not allow* for the $\epsilon_{i,s}$, but *does allow* for the differences between the observed right hand side variables and their expectations.

This means that when you are using discrete choice you should pay particular attention to; (i) the possibility of differences between observed values of right hand side variables and the perceptions that motivated the choice and to (ii) functional forms for the distribution of unobservables. If you are using this version of moment inequalities, you should pay particular attention to observable determinants of store characteristics; for example if the agent went to Whole Foods, you might want to look for a Whole Foods store (rather than any store) that is farther away (at least to see if that makes a difference).

Note. There is active research directed at maintaining the weaker assumptions of the moment inequality approach, and accommodating choice specific unobservables (see Dickstein and Morales, 2012, Pakes and Porter 2012.). It will clearly make the bounds wider but the extent of that is unclear. Indeed none of this will change the fact that the moment inequality methods will not give you a point estimate of a parameter, while standard discrete choice will. So if the standard discrete choice methods gives an acceptable approximation for the problem at hand, it is preferred.