The Diffusion of Wal-Mart and Economies of Density

Thomas J. Holmes*

Nov 2007 (still a rough draft)

Abstract

*The views expressed herein are solely those of the author and do not represent the views of the Federal Reserve Banks of Minneapolis or the Federal Reserve System. The research presented here was funded by NSF grant SES 0551062. I thank Junichi Suzuki, Julia Thornton, and David Molitor for excellent research assistance for this project. I thank Emek Basker for sharing data with me and I thank Ernest Berkas for help with the data.

1 Introduction

Wal-Mart opened its first store in 1962 and today there are over 3,000 Wal-Marts in the United States. The roll-out of stores displays a striking pattern that is best seen visually.[1] Wal-Mart started in a relatively central spot in the country (near Bentonville, Arkansas) and the store opening pattern radiated from the inside out. Wal-Mart never jumped to go to some isolated new location and then later fill out the area in between. Rather, the picture is one of a wave that gradually engulfs the whole the U.S. Wal-Mart always placed new stores close to where they already had store density.

This paper examines the benefits of such a strategy to Wal-Mart focusing on the logistic savings made possible by maintaining a dense network of stores. In particular it focuses on the advantages of this strategy for keeping stores near distribution centers. By maintaining short distances between a store and its distribution center, Wal-Mart of course saves on trucking costs. But there is an additional advantage in that it permits Wal-Mart to respond quickly to demand shocks. This is widely considered to be a key aspect of the Wal-Mart model. The aim of this paper is to quantify these gains to Wal-Mart.

A challenge in estimating these benefits is that Wal-Mart is notorious for being secretive—I am not going to get access to confidential data on its logistics costs. So it is not possible to conduct some kind of direct analysis relating costs to density. And even if Wal-Mart were to cooperate and make its data available, the benefits of being able to quickly respond to demand shocks might be difficult to quantify.[2] Instead, I pursue an indirect approach that exploits revealed preference. While density has a benefit, it also has a cost and I am able to make some progress in pinning down the cost. By examining Wal-Mart's choice behavior of how it trades off the benefit (not observed) versus the cost (observed with some work), it is possible to draw inferences about how Wal-Mart values the benefits.

The cost of high store density is that as stores get closer together, their market areas begin to overlap and new stores cannibalize sales from existing stores. The extent of such cannibalization is something I can estimate. To this issue, I bring store-level sales estimates from ACNeilsen and demographic data from the Census to estimate a model of demand for

[1] A video of Wal-Mart's store openings can be seen at www.econ.umn.edu/~holmes/research

[2] Of course, I would be delighted to have access to Wal-Mart's condifential data to give this a shot!

Wal-Mart at a rich level of geographic detail. I combine this with additional information about cannibalization that Wal-Mart itself releases in its annual reports. Using my sales model, I determine that Wal-Mart has encountered significant diminishing returns in sales as it has piled up many stores in the same area.

I write down a structural model of Wal-Mart and attempt to quantity parameters relating to density economies. Given the enormous number of different possible combinations of stores that can be opened, it is difficult to solve Wal-Mart's optimization problem. This makes conventional approaches used in the industrial organization literature infeasible. Instead, I use a perturbation approach. I consider a set of selected deviations from what Wal-Mart actually did and determine the set of parameters consistent with this decision. Using the procedure, I am able to determine a lower bound on the importance of density economies.

The paper contributes to the literature on entry and store location in retail. Related contributions include Bresnahan and Reiss (1991), Toivanen and Waterson (2005), Andrews et al (2004).

In addition to contributing to the literature on economies of density, the paper also contributes to a new and growing literature about Wal-Mart itself (e.g., Basker (forthcoming), Stone (1995), Hausman and Leibtag (2005), Ghemawat, Mark, and Bradley (2004)), Neumark et al (2005), Jia (2005). Wal-Mart has had a huge impact on the economy. It has been argued that this one company contributed a non-negligible portion of the aggregate productivity growth in recent years. Wal-Mart is responsible for major changes in the structure of industry, of production, and in of labor markets. One good question is: what exactly is a Wal-Mart, why is it different from a K-Mart or a Sears? One thing that distinguishes Wal-Mart is its emphasis on logistics and distribution. (See, for example, Holmes (2001)). It is plausible that Wal-Mart's recognition of economies of density and its knowledge of how to exploit these economies distinguished it from K-Mart and Sears and is part of the secret of Wal-Mart's success.

2 Model

A retailer (Wal-Mart) has a network of stores supported by a network of distribution centers. The model specifies how Wal-Mart's revenues and costs in a period depend on the configuration of stores and distribution centers that are open in the period. And the model specifies how the networks change over time.

There are two categories of merchandise: *general merchandise* (abbreviated by g) and *food* (abbreviated by f). There are two kinds of Wal-Mart stores. A *regular store* sells only general merchandise. A *supercenter* sells both general merchandise and food.

There are a finite set $\mathcal{L}$ of locations in the economy. Locations are indexed by $\ell = 1, ...L$. Let $d_{\ell\ell'}$ denote the distance in miles between any given pair of locations ℓ and ℓ'. At any given period t, a subset $B_t^{Wal} \subseteq \mathcal{L}$ of locations have a Wal-Mart. Of these, a subset $B_t^{Super} \subseteq B_t^{Wal}$ are supercenters and the rest are regular stores. In general the number of locations with Wal-Marts will be small relative to the total set of locations and a typical Wal-Mart will draw sales from many locations.

Sales revenues at a particular store depend upon the store's location and its proximity to other Wal-Marts. Let $R_{jt}^{g}(B_t^{Wal})$ be general merchandise sales revenue of store j at time t given the set of Wal-Mart stores open at time t. If store j is a supercenter, then its food sales $R_{jt}^{f}(B_t^{Super})$ analogously depend upon the configuration of supercenters. The model of consumer choice will be specified below in Section 4 from which this demand function will be derived. In this demand structure, Wal-Mart stores that are near each other will be regarded as substitutes by consumers. That is, increasing the number of nearby stores will decrease sales at a particular Wal-Mart.

I abstract from price variation and assume Wal-Mart sets constant prices across all stores and over time. In reality, prices are not always constant across Wal-Marts, but the company's Every Day Low Price (EDLP) policy makes this a better approximation for Wal-Mart than it would be for many retailers. Let μ^g denote the gross margin on general merchandise. Thus for store j at time t, $\mu^g R_{jt}^{g}(B_t^{Wal})$ is sales receipts less cost of goods sold for general merchandise. Analogously, μ^f is the gross margin on food.

In the analysis there will be three components of cost that will be relevant besides cost of goods sold: (1) distribution costs, (2) variable store costs, and (3) fixed costs at the store level. I describe each in turn.

Distribution Cost

Each store requires distribution services. General merchandise is supplied by a Regional Distribution Center (RDC) and food by a Food Distribution Center (FDC). For each store, these services are supplied by the closest distribution center. Let $d^g_{j,t}$ be the distance in miles from store j to the closest RDC at time j and analogously define d^f_j. If store j is a supercenter, its distribution cost in time t is

$$Distribution\_Cost_{jt} = \tau d^g_{jt} + \tau d^f_{jt}.$$

where the parameter τ is the cost per mile per period per merchandise segment (general or food) of servicing this store. If j carries only general merchandise the cost is τd^g_j.

Note that the distribution cost here is a fixed cost that does not depend upon the volume of store sales. This would be an appropriate assumption if Wal-Mart made a single delivery run from the distribution center to the store each day. The driver's time is a fixed cost and the implicit rental on the tractor is a fixed cost that must be incurred regardless of the size of the load. To keep a tight rein on inventory and to allow for quick response, Wal-Mart aims to have daily deliveries even for its smaller stores. So there clearly is an important fixed cost component to distribution. Undoubtedly there is a variable cost component as well but for simplicity I abstract from it.

Variable Costs

The larger the sales volume at a store, the greater the number of workers needed to staff the checkout lines, the larger the parking lot, the larger the required shelf space, and the bigger the building. All of these costs are treated as variable in this analysis. It may seem odd to treat building size and shelving as a variable input. However, Wal-Mart very frequently updates and expands its stores. So in practice, store size is not a permanent decision that is made once and for all but is rather a decision made at multiple points over time. Treating store size as a variable input simplifies the analysis significantly.

Assume that the variable input requirements at store j are all proportionate to sales volume R_j,

$$\begin{aligned} Labor_j &= \nu^{Labor} R_j \\ Land_j &= \nu^{Land} R_j \\ OtherCosts_j &= \nu^{Other} R_j. \end{aligned}$$

Wages and land prices vary across locations and across time. So let $w_{j,t}$ and $r_{j,t}$ denote the wage and land rental rate that store j faces at time t. Other costs consists of all other costs that vary with sales, including the rental rate on structure and equipment in the store (the shelving, the cash registers, etc.) The other cost component of variable costs is assumed to be the same across locations and the price is normalized to one.

Fixed Cost

We might expect there to be a fixed cost with operating a store. To the extent the fixed cost is the same across locations, it will play no role in the analysis of where Wal-Mart places it stores. We are only interested in the component of fixed cost that varies across locations.

From Wal-Mart's perspective, urban locations have some disadvantages compared to non-urban locations. Land rents and wages are higher in urban areas and these have already been taken into account above. But disadvantages go beyond land rent and wages. The Wal-Mart model of a big box store at a convenient highway exit is not applicable in a very urban location. Moreover, Sam Walton was very concerned about the labor force to be found in urban locations, as he explained in his autobiography (Walton (1992)). We might expect, for example, that urban workers would be less interested in joining on the trademark Wal-Mart cheer (Give me a "W"...). Urban locations are more susceptible to unions and Wal-Mart has been very up front about not wanting unions in its stores.

To capture potential disadvantages of urban locations, the fixed cost of operating store j is written as a function $f(m_j)$ of the population density m_j of the store's location. The funcational form is quadratic in logs of a truncated population density measure,

$$f(m_j) = \omega_0 + \omega_1 \ln(m_j) + \omega_2 \ln(m_j)^2 \tag{1}$$

A supercenter is actually two stores, a general merchandise and food store, so the fixed cost is paid twice. It will be with no loss of generality in our analysis to assume that the constant term $\omega_0 = 0$ since the only component of the fixed cost that will matter in the analysis is the part that varies across locations.

2.0.1 Dynamics

Everything that has been discussed so far considers quantities for a particular time period. I turn now the dynamics aspects of the model. Wal-Mart operates in a deterministic environment in discrete time where it has perfect foresight. The general problem Wal-Mart faces is to determine for each period:

1. How many new Wal-Marts and how many new supercenters to open?
2. Where to put the new Wal-Marts and supercenters?
3. How many new distribution centers to open?
4. Where to put the new distribution centers?

The main focus of the paper is on part 2 of Wal-Mart's problem. The analysis conditions on the answers to 1, 2, and 4, in terms of what Wal-Mart actually did, and solves Wal-Mart's problem of getting 2 right. Of course, if Wal-Mart's actual behavior solves the true problem of choosing 1 through 4, then it also solves the constrained problem of choosing 2, conditioned on 1, 3, and 4 being what Wal-Mart actually did.

Getting at part 1 of Wal-Mart's problem—how many new stores Wal-Mart opens in a given year—is far afield from the main issues of this paper. In its first few years, Wal-Mart added only one or two stores a year. The number of new store openings has grown substantially over time and in recent years they sometimes number several stores in one week. Presumably capital market considerations have played an important role here. This is an interesting issue, but not one I will have anything to say about with this project.

Problems 3 and 4 regarding distribution centers are closely related to the main issue of this paper. I will have something to say about this later in the paper.

Now for more notation. To begin with, the discount factor each period is β. The periods will be a year and β will be set to $\beta = .95$.

As defined earlier, B_t^{Wal} is the set of Wal-Mart stores in period t and B_t^{Super} is the subset of supercenters. Assume that once a store is opened, it never shuts down. This assumption simplifies the analysis considerably and is not inconsistent with Wal-Mart's behavior as it rarely closes stores.[3] Then we can write $B_t^{Wal} = B_{t-1}^{Wal} + A_t^{Wal}$, where A_t^{Wal} is the set of new stores opened in period t. Analogously, a supercenter is an absorbing state, $B_t^{Super} = B_{t-1}^{Super} + A_t^{Super}$, for A_t^{Super} the set of new supercenter openings in period t. A supercenter can open two ways. It can be a new Wal-Mart store that opens as a supercenter as well. Or it can be a conversion of an existing Wal-Mart store.

Let N_t^{Wal} and N_t^{Super} be the number of new Wal-Marts and supercenters opened at t, i.e. the cardinality of the sets A_t^{Wal} and A_t^{Super}. Choosing these values was defined as part 1 of Wal-Mart's problem. These are taken as given here. Also taken as given is the location of distribution centers of each type and their opening dates.

There is exogenous productivity growth of Wal-Mart according to a growth factor ρ_t in period t. If Wal-Mart where to hold fixed the set of stores and demographics also stayed the same, from period $t-1$ to period t, then revenue and all components of costs would grow at the factor ρ_t. As will be discussed later, the growth of sales per store of Wal-Mart has been remarkable. Part of this growth is due the gradual expansion of its product line, from hardware and variety items to eye glasses and tires. The part of this growth that is due to expansion to supercenters is explicitly accounted for here. The residual part is not modeled explicitly and but is taken as exogenous instead. The role the parameter ρ_t plays in the analysis is like a discount factor.

A policy choice of Wal-Mart is a vector $a = (A_1^{Wal}, A_1^{Super}, A_2^{Wal}, A_2^{Super}, ..., A_T^{Wal}, A_T^{Super})$ specifying the new stores opened in each period t. Define a choice vector a to be feasible if the number of store openings openings in period t under policy a equals what Wal-Mart actually did, i.e. N_t^{Wal} new stores in a period and N_t^{Super} supercenter openings. Wal-Mart's problem at time 0 is to pick a feasible a to maximize

$$\max_a \sum_{t=1}^{T} (\rho_t \beta)^{t-1} \left[\sum_{j \in B_t^{Wal}} \left[\pi_{jt}^g - f_{jt}^g - \tau d_{jt}^g \right] + \sum_{j \in B_t^{Super}} \left[\pi_{jt}^f - f_{jt}^f - \tau d_{jt}^f \right] \right]. \qquad (2)$$

[3] Footnote to back this up.

where the operating profit for merchandise segment $k \in \{g, f\}$ at store j in time t is

$$\pi_{jt}^{k} = \mu^{k} R_{jt}^{k} - w_{jt} Labor_{jt}^{k} - r_{jt} Land_{jt}^{k} - Other_{jt}^{k}$$

and where d_{jt}^{k} is the distance to the closest type k distribution center at time t. Again, the sequence of distribution center openings is taken as given.

3 The Data

There are five main data elements used in the analysis. The first element is store-level data on sales and other store characteristics. The second is opening dates for stores, supercenters, general distribution centers and food distribution centers. The third is demographic information from the Census. The fourth element is data on wages and rents across locations. The fifth is other information about Wal-Mart from annual reports.

Data element one, store-level data variables such as sales, was obtained from *TradeDimensions*, a unit of ACNeilsen. This data provides estimates of store-level sales for all Wal-Marts open as of the end of 2005. This data is the best available and is the primary source of market share data used in the retail industry. Ellickson (2007) is a recent user of this data for the supermarket industry.

Table 1 presents summary statistics of annual store-level sales and employment for the 3,176 Wal-Marts in existence in the contiguous part of the United States as the end of 2005.[4] (Alaska and Hawaii are excluded in all of the analysis.) As of this point in time, almost two thirds of all Wal-Marts (1,980 out of 3,176) are supercenters. The average Wal-Mart racks up annual sales of \$70 million. The breakdown is \$47 million per regular store and \$85 million per supercenter. The average employment is 255 employees.

The second data element is opening dates of Wal-Mart facilities. The appendix provides the details of how this information was collected. Briefly, exact opening dates for new Wal-Mart stores were obtained from information posted by Wal-Mart on its website. Many of Wal-Mart's supercenters are conversions of Wal-Marts that previously existed as regular

[4] The Wal-Mart Corporation has other types of stores that I exclude in the analysis. In particualr, I am excluding Sam's Club (a wholesale club) and Neighborhood Market stores, Wal-Marts recent entry into the pure grocery store segment.

stores. For conversion dates 2001 and after I use data from Wal-Mart's web site, but for dating conversions before this I am grateful for data Emek Basker. Opening dates for regional distribution centers and food distribution centers were compiled from various sources such as Lexis-Nexis and the web. Table 2 tabulates opening dates for the four types of facilities by decade.

The third data element, demographic information, comes from the three decennial censuses, 1980, 1990, and 2000. The data is at the level of the *block group*, a geographic unit finer than the Census tract. Summary statistics are provided in Table 3. In 2000, there were 206,960 block groups with an average population of 1,350. The Census provides information about the geographic coordinates of each block group which I use extensively in the analysis. For each block group I determine all the block groups within a five mile radius and add up the population of these neighboring areas. Population within a five mile radius is the density measure m used in the analysis. With density defined this way, the mean density in 2000 across blockgroups was 219,000 people within a five-mile radius. The table also reports mean levels of per capita income, share old (65 or older), share young (21 or younger), and share black. The per capita income figure is in 2000 dollars for all the Census years using the CPI as the deflator.[5]

The fourth data element is information about local wages and local rents. The wage measure is average retail wage by county from County Business Patterns. This is payroll divided by employment. I use annual data over the period 1977 to 2004. It is difficult to obtain a consistent measure of land rents over a detailed cross-section and over a long period of time. There exist commercial rent data at the metropolitan level but information on rents within metropolitan areas is needed here. To proxy land rents, I use information about residential property values from the 1980, 1990, and 2000 decennial censuses. For each Census year and each store location, I create an index of property values by adding up the total value of residential property within a two miles of the store's location and scaling it so the units are in inflation adjusted dollars per acre. See the appendix for how the index is constructed. To obtain values between Census years, I interpolate. I supplement the Census data with property tax data on property valuations of actual Wal-Mart store locations in

[5] Per capita income is truncated from below at $5,000 in year 2000 dollars.

Iowa and Minnesota. As discussed in the appendix, there is a high correlation between the tax assessment property valuations of a Wal-Mart site and the property value index.

The fifth data element is information from Wal-Mart's annual reports including information about aggregate sales for earlier years. (TradeDimensions provides only current data.) I also use information provided in the "Management Discussion" section of the reports on the degree to which new stores cannibalize sales of existing stores. The specifics of this information is explained below when the information is incorporated into the estimation.

4 Estimates of Operating Profits

This section estimates the components of Wal-Mart's operating profits. Part 1 specifies the demand model and Part 2 estimates it. Part 3 treats various cost parameters. The estimates of demand and costs employ data from 2005. Part 4 explains how the estimates for 2005 are extrapolated to other years.

4.1 Demand Specification

Conditioned up shopping at some Wal-Mart, presumably a consumer will tend shop at the Wal-Mart closest to home. Nevertheless, for various reasons, some consumers will shop at other Wal-Marts. For example, a particular consumer may pass a Wal-Mart on the way to work and it may be more convenient to shop there rather that at the Wal-Mart closest to home. To allow for substitution possibilities such as this, I specify a model in which the various Wal-Marts in the vicinity of a consumer are differentated products that the consumer chooses between. Following common practice in the literature, I employ a discrete choice approach.

For a given location ℓ, let n_ℓ denote the *population* of location ℓ and let m_ℓ be the *population density* at ℓ. In the empirical work a location will be a Census Block Group and the population density measure will be the number of people residing within a five mile radius of the Block Group.

Let $y_{\ell j}$ denote the distance in miles between location ℓ and store location j. Define $\bar{B}_\ell^{Wal}$ to be the set of Wal-Marts in the *vicinity* of the consumer's location, defined to be those locations within 25 miles,

$$\bar{B}_{\ell}^{Wal} = \left\{ j, j \in B^{Wal} \text{ and } y_{\ell j} \leq 25 \right\}.$$

(The time subscript t is left out here because the time period is held fixed in this part of the paper.) Analogously define $\bar{B}_{\ell}^{Super}$

In the model, the shopping decision for general merchandise is separated from the shopping decision for food. In general, we expect that there would be some complementarity—once a consumer is in the store to buy a lawn mower, the consumer might also buy meat for dinner. I ignore this issue in part because I don't have good data to get a handle on the issue. (In particular, I don't have a clean breakdown between the two segments in my store-level data.) I have no reason to believe that abstracting from such a complementarity biases my results in a particular direction.

I explain the purchase decision for general merchandise; the food purchase choice problem is analogous. Consider a consumer at a particular location ℓ. The consumer has a budget λ^g for spending on general merchandise. The consumer makes a discrete choice between buying general merchandise from the outside alternative (labeled $j = 0$) or from one of the Wal-Marts in $\bar{B}_{\ell}^{wal}$ (assuming $\bar{B}_{\ell}^{wal}$ is nonempty).

If the consumer chooses the outside alternative 0, utility is

$$u_0 = b(m_\ell) + z_\ell \omega + \varepsilon_0. \tag{3}$$

The first term is a function $b(\cdot)$ that depends upon the population density m_ℓ at the consumer's location ℓ. Assume $b'(m) \geq 0$; i.e., the outside option is better in higher population density areas. This is a sensible assumption as we would expect there to be more substitutes for a Wal-Mart in larger markets for the usual reasons. A richer model of demand would explicitly specify the alternative shopping options available to the consumer. I don't have sufficient data to conduct such an analysis so instead specify the reduced-form relationship between $b(m_\ell)$ and population density. The functional form used in the estimation is

$$b(m) = \omega_0 + \omega_1 \ln(\underline{m}) + \omega_2 \left(\ln(\underline{m})\right)^2$$

where

$$\underline{m}_j = \max\left\{1, m_j\right\}. \tag{4}$$

The units of the density measure is thousands of people within a five mile radius. By truncating m at one, $\ln(m)$ is truncated at zero. All locations with less then one thousand people within five miles are grouped together.[6]

The second term of (3) allows demand for the outside good to depend upon a vector z_ℓ of location characteristics that impact utility through the parameter vector ω. In the empirical analysis, these characteristics will include the demographic characteristics of the block group and income.

The third term is a logit error term. Assume this is drawn i.i.d. across all consumers living in the block group ℓ.

This explains utility of the outside alternative. Now consider the utility from buying at Wal-Mart. The utility from buying at a particular Wal-Mart $j \in \bar{B}_\ell^{Wal}$ is

$$u_{\ell j} = -h\left(m_\ell\right) y_{\ell j} + x_j \gamma + \varepsilon_j. \tag{5}$$

for $h(m)$ parameterized by

$$h(m) = \delta_0 + \delta_1 \ln(\underline{m}).$$

The first term of (5) is the disutility of commuting $y_{\ell j}$ miles to the store from the consumer's home. The coefficient $h\left(m_\ell\right)$ can be interpreted as a transportation cost per mile. The specification allows the transportation cost to depend upon population density. The second term of (5) allows utility to depend upon other characteristics x_j of Wal-Mart store j. The other store characteristic included in the empirical analysis is store age. In this way, it is possible in the demand model for a new store to have less sales, everything else the same. This captures in a crude way that it takes a while for a new store to ramp up sales. The last term is the logit error ε_j. A consumer who finds store j a convenient place to stop on the way home from work can be interpreted as a consumer with a high value of ε_j.

Using the standard logit formulas, the probability that a consumer at location ℓ find's Wal-Mart j to be be the best option is

$$p_{j\ell}^g = \frac{\exp(\delta_{j\ell})}{\left[\exp\left(\delta_{0\ell}\right)\right] + \sum_{k \in \bar{B}_\ell^{Wal}} \exp(\delta_{k\ell})} \tag{6}$$

[6] This same truncation is applied throughout the paper.

for

$$\delta_{0\ell} \equiv b(m_\ell) + z_\ell \omega$$
$$\delta_{j\ell} \equiv -h(m_\ell) y_{\ell j} + x_j \gamma.$$

The model's predicted general merchandise revenue for store j is

$$R_j^g = \sum_{\{\ell | j \in \bar{B}_\ell^g\}} \lambda^g \times p_{j\ell}^g \times n_\ell. \tag{7}$$

In words, there are n_ℓ consumers at location ℓ and a fraction $p_{j\ell}^g$ of them are shopping at j where they will each spend λ^g dollars.

Spending on food is modeled the same way. The parameters are the same except for the spending λ^{groc} per consumer. The formula for food revenue R_j^f at store j is analogous to (7). Even thought the parameters for food are the same as for general merchandise, the probability $p_{j\ell}^f$ a consumer at ℓ shops at j for food will differ from the probability $p_{j\ell}^g$ the consumer shops for general merchandise. This follows because the set of alternatives for food $\bar{B}_\ell^f$ is in general different from the set of alternatives $\bar{B}_\ell^g$ for general merchandise.

4.2 Demand Estimation

Given a vector θ of parameters from the demand model, we can plug in the demographic data and obtain predicted values of general merchandise sales $\hat{R}_j^g(\theta)$ for each store j from equation (7) and predicted values of food sales $\hat{R}_j^f(\theta)$.

The data has *all commodity sales volume* for each store. Call this R_j. General merchandise is all items sold at a regular Wal-Mart. (So beverages like Coke and Pepsi sold at regular Wal-Marts are considered general merchandise.) So for regular stores, $R_j = R_j^g$, by definition. For supercenters, all commodity sales volume includes general merchandise and food, $R_j = R_j^g + R_j^f$.

Let η_j be the difference between log actual sales and log predicted sales for store j. For regular stores this is

$$\eta_j^{Wal} = \ln(R_j) - \ln(\hat{R}_j^g(\theta)).$$

For supercenters, this is

$$\eta_j^{Super} = \ln(R_j) - \ln(\hat{R}_j^g(\theta) + \hat{R}_j^f(\theta)).$$

Assume the discrepancies η_j^{Wal} and η_j^{Super} are i.i.d. normally distributed measurement error. (The store-level sales figures in the TradeDimensions data are estimates so certainly measurement error is issue.) The model is estimated using maximum likelihood and the coefficients are reported in Table 4 in the column label "Unconstrained Demand Model."

The extent that new stores cannibalize the sales of existing stores will make a big difference in the subsequent analysis. So our first order of business is to assess how well the demand model is doing in getting this right. Fortunately, Wal-Mart has provided information that is helpful in this regard. Wal-Mart's annual report for 2004 disclosed,

> "As we continue to add new stores in the United States, we do so with an understanding that additional stores may take sales away from existing units. We estimate that comparative store sales in fiscal year 2004, 2003, 2002 were negatively impacted by the opening of new stores by approximately 1%."

This same paragraph was repeated in the 2006 annual report with regards to fiscal year 2005 and 2006. This information is summarized in Table 5. [7]

For each vector θ of model parameters, calculate what sales would have been in a particular year to preexisting stores if no new stores had opened in the year and if there were no new supercenter conversions. Next calculate predicted sales to pre-existing stores when the new store openings and supercenter conversions for particular year take place. Define the percentage difference to be the *cannibalization rate* for that year. This is the model analog of what Wal-Mart is disclosing.

Table 5 reports the cannibalization rate for various years using the estimated demand model. The parameter vector is the same across years. What varies over time are the new stores, the set of pre-existing stores and the demographic variables.[8] The demand model—estimated entirely off of the 2005 cross-section store-level sales data—does a very good fitting the cannibalization rates reported by Wal-Mart. For the years that Wal-Mart disclosed that the rate was "approximately one", the estimated rates range from .67 to 1.43.

[7] Wal-Marts fiscal year ends January 31. So the fiscal year corresponds (approximately) to the previous calendar year. For example, the 2006 fiscal year corresponds to the 2005 calendar year. Thoughout the paper we use the calendar year, not the fiscal year.

[8] To obtain demographic characteristics between Census years, I interpolate as discussed below.

It is interesting to note that sharp increase in the estimated cannibalization rate beginning in 2002. Evidently, Wal-Mart reached some kind of saturation point in 2001. Given the pattern in Table 5, it is understandable that Wal-Mart has felt the need to disclose the extent of cannibalization in recent years.

In what follows, the estimated upper bound on the degree of density economies will be closely connected to the degree of cannibalization. The more cannibalization Wal-Mart is willing to tolerate, the higher the inferred density economies. The estimated cannibalization rates of 1.38, 1.43, and 1.27 for 2003, 2004, 2005 are certainly "approximately one" but one may worry that these rates are on the high end of what would be consistent with Wal-Mart's reports. To explore this issue further, I estimate a second demand model where the cannibalization rate for 2005 is constrained to be exactly one. The estimates are reported in the final column of Table 4. The goodness of fit under the constraint is close to the unconstrained model, although a likelihood ratio test leads to a rejection of the constraint. In the interests of being conservative in my estimate of an upper bound on density economies, I will use the constrained model through as the baseline model.

A few remarks about the parameter estimates. Recall that λ^g and λ^f are spending per consumer in the general merchandise and food categories. The estimates can be compared to aggregate statistics. For 2005, per capita spending in the U.S. was 1.8 in general merchandise stores (NAICS 452) and 1.8 in food and beverage stores (NAICS 445) (in thousands of dollars). The aggregate statistics match well the model estimates ($\lambda^g = 1.9$ and $\lambda^f = 1.9$ in the constrained model, $\lambda^g = 1.7$, $\lambda^f = 1.6$ in the unconstrained model), though we would not expect them to match exactly.[9]

The parameter estimates reveal that, as hypothesized, the outside good is better in more dense areas and that utility decreases in distance travelled to a Wal-Mart. To get a sense of the magnitudes, Table 6 examines how predicted demand in a block group varies with

[9] On one hand, the general merchandise category includes Saks Fifth Avenue which is not likely to be in the same spending budget with a Wal-Mart. One the other hand, the general merchandise category does not include the electronics giant Best Buy; a large portion of this merchandise would be in the same spending budget with Wal-Mart. Both of these categories are relatively small (electronics is less than a fifth of general merchandise sales) so perhaps it is not a surprise that my estimate of λ^g is so close to U.S. per capita spending in this category.

population density and distance to the closest Wal-Mart. (The table is constructed with the constrained model but things look very similar with the unconstrained model.) The table reports the probability that a consumer shops at Wal-Mart for general merchandise. For the analysis, the demographic variables in Table 3 are set to their mean values. There is assumed to be only one Wal-Mart (two or more years old) in the vicinity of the consumer (i.e. within 25 miles) and the distance to this Wal-Mart is varied. Consider the first row, where distance is set to zero (the consumer is right-next door to a Wal-Mart) and population density is varied. The negative effect of population density on demand is substantial. A rural consumer right next to a Wal-Mart shops there with a probability that is essentially one. With a population density of 50 this falls to .72 and at 250 it falls to only .24. In a large market there are many substitutes; even a customer right next to a Wal-Mart is not likely to shop there. While *per capita* demand falls, overall demand overwhelmingly increases in large markets. A market that is 250 times as large as an isolated market may have a per capita demand only a fifth as large, but overall demand is 50 times as large.

Next consider the effect of distance holding fixed population density. In a very rural area, increasing distance from 0 to 5 miles has only a small effect on demand. This is exactly what we would expect. Raising the distance further from 5 to 10 miles has an appreciable effect, .98 to .57, but still much demand remains. Contrast this with larger markets. For market size 250, an increase in distance from 0 to 5 miles reduces demand by on the order of 80 percent while the effect of distance in rural markets is miniscule. This is what we would expect.

Demand varies by demographic characteristics in interesting ways. Wal-Mart is an inferior good in that demand decreases in income. (This is the same thing as saying the coefficient on per capita income for the outside good is positive.) Demand for Wal-Marts is lower among blacks and young people.

The only store characteristic used in the demand model (besides location) is store age. This is captured with a dummy variable for stores that have been open two or more years. This variable enters positively in demand. So everything else the same, older stores attract more sales.

4.3 Variable Costs at the Store Level

In the description of the model, the required labor input at the store level was assumed to be proportionate to sales. To get a sense of the plausibility of this assumption, Figure 1 provides a scatter plot of square footage of each store along with sales per employee of each store in 2005. Also plotted are the fitted values of a locally weighted regression. At the bottom end of the size distribution, there is evidence of increasing returns. But things flatten out and there is roughly constant returns over most of the store size distribution. The weighted average over all stores is \$277 thousand dollars in sales per employee. Equivalently, there are 3.61 store employees per million dollars of annual sales. I use this as the estimate of the fixed labor coefficient, $\nu_{Labor} = 3.61$. To covert this into a cost of labor at a particular store, the coefficient is multiplied by average retail wage (annual payroll per worker) in the county where the store is located. Table 7 reports information about the distribution of labor costs across the 2005 set of Wal-Mart stores. The median store faces a labor cost of \$20,700 per worker. Given $\nu_{Labor} = 3.61$, this translates into a labor cost of $3.61 \times 20,700$ per million in sales or equivalently 7.5 percent of sales. The highest labor costs can be found at stores in San Jose, California where wages are almost twice as high as they are for the median store.

An issue that needs to be raised about the County Business Patterns wage data is measurement error. Dividing annual payroll by employment is a crude way to measure labor costs because it doesn't take into account potential variations in hours per worker (e.g. part-time versus full-time) or potential variations in labor quality. The empirical procedure used below explicitly takes into account this measurement error.

Turning now to land costs, the appendix describes the construction a *property value index* for each store. It is an estimate of the average value of residential property per acre within a two-mile radius of each store. The decennial censuses are used to construct this measure for each store location. For 46 Wal-Mart locations in Minnesota and Iowa, actual land valuations for the store properties were obtained from property tax records. Define the *land-value/sales ratio* to be land value from the tax records as a percent of the (fitted) value of 2005 sales for each store. Figure 2 is a scatter plot of the land-value/sales ratio and the property-tax index for the 46 stores with the available information. Consider for example, the circled observation with the highest value of the index. This is a store location in the city of St. Paul where the index takes a value of $465,000 per acre. The land for the store site is valued at six million dollars which is about 12 percent of the 2005 sales of the store. Figure 2 exhibits a strong connection between the index and land-value/sales ratio. Stores with very low values of the index are in rural locations with very low land-value/sales ratios. Fitting a line through the origin yields a slope of .036 and a standard error of .003. I use this regression line to obtain fitted values of the land-value/sales ratio for all Wal-Mart stores. The bottom panel of Table 7 reports the distribution of the estimated land-value sales ratio across stores. The highest value is 65 percent of sales attained at a store in Silicon Valley in California. While is sensible for such a place to end up as the location with the highest land prices, there is clearly measurement error with this method of estimating land prices. The empirical procedure below explicitly takes this measurement error into account.

The price of land of course varies substantially across locations. The cost of structures (the cinderblocks used to make the wall, the shelving, etc.) varies much less. (See a further discussion in the appendix.) In what follows I calculate the annual rent to be 20 percent of the estimated land value of the site. So for the median store from Table 7 (the Wal-Mart in Cleburne, Texas), this implies a annual rental rate of about half a percent of sales ($.5 \approx .2 \times 2.4$). This would be on the low side of what we might expect the rent to be for both land *and* structures. But what I am trying to do here is account for costs that would be different *across* locations. The rent on the cinderblock walls, the shelving and the asphalt parking lots will be approximately the same across locations and can be folded into a component of variable cost that is common across locations. Variable costs that are common across all stores is the issue to which we now turn.

Wal-Mart's gross margin over the years has ranged from .22 to .26. (See Wal-Marts annual reports.) To be consistent with this, the gross margin is set equal to $\mu = .24$.

Wal-Mart has reported over the years *operating selling, general and administrative expenses* that are in the range of 16 to 18 percent of sales. Included in this is the store-level labor cost discussed above that is on the order of 7 percent of sales and has already been taken account of. Also included in this cost is the cost of running the distribution sytem, the fixed cost of running central administration and other costs I don't want to include as variable costs. For the baseline analysis, I set the residual variable cost parameter $\nu^{other} = .07$. Netting this out of the gross margin μ yields a net margin $\mu - \nu^{other} = .17$. In the analysis, the breakdown between μ and ν^{other} is irrelevant, only the difference. In what follows, I use the .17 percent figure as a baseline and experiment to see how results change when it varies.

The analysis so far has explained how to calculate the store-level operating profit of store j in 2005 as

$$\pi_{j,2005} = \left(\mu - \nu^{other}\right)\left(R^g_{j,2005} + R^f_{j,2005}\right) - LaborCost_{j,2005} - Landrent_{j,2005} \tag{8}$$

where the sales revenue comes from the 2005 demand model and labor cost and land rent are explained above. The next step is to extrapolate this model to earlier years.

4.4 Extrapolation to Other Years

We have a demand model for 2005 in hand but need models for earlier years, as well. To get them, assume demand in earlier years is the same as in 2005 except for the multiplicative scaling factorρ_t introduced above in the definition of Wal-Mart's problem (2). For example, the 2005 demand model with no rescaling predicts that, at the 1972 store set and 1972 demographic variables, average sales per store (in 2005 dollars) is \$31.5 million. Actual sales per store (in 2005 dollars) for 1971 is \$7.4 million. The scale factor for 1971 adjusts demand proportionately so that the model exactly matches 1971 sales. Over the1971 to 2005 period, this corresponds to a compound annual real growth rate of 4.4 percent. Wal-Mart significantly widened the range of products it sold over this period (to include tires, eyeglasses, etc.). The growth factor is meant to capture this. The growth factor calculated in this manner has levelled off in recent years to around one percent a year. Wal-Mart has been expanding by converting regular stores to supercenters. That kind of expansion is captured explicitly in the model rather than indirectly though exogenous growth.

Demographics change over time and this is taken into account. For the 1980, 1990, and 2000, I take the decennial census for that year.[10] For years in between I use a convex combination of the censuses. For example, for 1984 I convexify by placing .6 weight on 1980 and .4 weight on 1990. More specifically, 60 percent of the people in the 1980 block group are

[10] I use 1980 for years before 1980 and 2000 for years after 2000.

assumed to be still around as potential Wal-Mart customers and 40 percent of the 1990 block group consumers have already arrived as of 1984. This procedure keeps the geography clean, since the issue of how to link block groups over time is avoided. (Blockgroup definitions do not stay constant across census years.)

5 Preliminary Evidence of a Tradeoff

This section provides some preliminary evidence of an economically significant tradeoff to Wal-Mart. Namely, the benefits of increased economies of density come at the cost of more cannibalization of existing stores. This section puts to work the demand model and other components of operating profits compiled above.

Consider some Wal-Mart store j that first opens in time t. Define the *incremental sales* $R_{j,t}^{k,inc}$ of store j to be what the store adds to total Wal-Mart sales in segment $k \in \{g, f\}$ in its opening year t, relative to what sales would otherwise be across all other stores open that year. The incremental sales of store #1 opening in 1962 equal the sales $R_{j,1962}^{k}$ that year. For later stores, however, incremental sales are in general less than store sales, $R_{j,t}^{k,inc} \leq R_{j,t}^{k}$, because some part of the sales may be diverted from other stores. Using the demand model, we can calculate $R_{j,t}^{k,inc}$ for each store.

Table 8 reports that the average annual incremental sales in general merchandise across all Wal-Mart stores in the year the stores opened is \$36.3 million (in 2005 dollars throughout). Analogously, average incremental sales in food from new supercenters is \$40.2 million. The growth factor adjustment is not used here to make store openings across different years comparable. (So the 2005 demand model is applied to the store configurations and demographics of the earlier years but with no multiplicative scale adjustment ρ_t.) In an analogous manner, we can use (8), to determine the *incremental operating profit* of each store at the time it opens. The average incremental profit in general merchandise from a new Wal-Mart is \$3.1 million and in food from new supercenters is \$3.6 million. Finally, we can ask how far a store is from the closest distribution center in the year it is opened. On average a new Wal-Mart is 168.9 miles from the closest regular distribution center when it opens and a new supercenter is 137.0 miles from the closest food distribution center.

Incremental sales and operating profit can be compared to what sales and operating profit would be if a new store were a stand-alone operation. That is, what would sales and operating profits be at the store if it were isolated so that none of its sales are diverted to or from other Wal-Mart stores in the vicinity? Table 8 shows for the average new Wal-Mart, there is a big difference between stand-alone and incremental values, implying a substantial degree of market overlap with existing stores. Average stand-alone sales is $41.4 million compared to an incremental value of $36.3 million, approximately a 10 percent difference. Two considerations make it clear that this big difference is not inconsistent with one percent cannibalization rates in the model reported earlier in Table 5. First, stand-alone sales include sales that the new store never gets because they remain in the existing stores (but would diverted to the new store if existing stores shut down.) Second, the denominator of the cannibalization rate from Table 5 includes all pre-existing stores, including those areas of the country were Wal-Mart is not adding any new stores. Taking an average over the country as a whole understates the degree of cannibalization taking place where Wal-Mart is adding new stores.

Define the *Wal-Mart Age* of a state to be the number of years that Wal-Mart has been in the state. [11] The remaining rows in Table 8 classify stores the Wal-Mart age of their state at the store's opening. Those store in the row labeled "'1-2" are the first stores opened in the states where they are located. Those stores in the row labeled 21 and above are opened when Wal-Mart has been in the state for over 20 years.

Table 8 shows that incremental operating profit in a state falls over time as Wal-Mart adds stores in a state and the market areas increasingly begin to overlap. Things are actually flat the first five years at 3.5 million in incremental operating profit for general merchandise. But it falls to 3.3 million in the second five years and then to 2.9 million and lower beyond that. An analogous pattern holds for food. This pattern is a kind of diminishing returns. Wal-Mart is getting less incremental operating profit from the later stores it opens in a state. The table also reveals a benefit from opening stores in states where Wal-Mart has had a long presence. The incremental distribution center distance falls substantially as we move down

[11] For the purposes of this analysis, the New England states are treated as a single state. Maryland, Delaware and the District of Columbia are also aggregated.

the table. The very first stores in a state on average on the order of 300 miles from the closest distribution center. Stores opening in state where Wal-Mart has been there over 20 years are less than 100 miles from the closest distribution center. There is a tradeoff here: the later stores deliver lower operating profit but can be supplied cheaper.

As argued in Section 2, the fixed cost of operating a store may be higher in high density areas. The comparisons in Table 8 control for variable costs across locations but not fixed costs that depend upon density. Table 9 runs the regression analog of Table 8 with a control for population density that is quadratic in logs, following the specification of the fixed cost (1). In addition, fixed state effects are included in the regression. The idea is to hold fixed population density and determine how incremental profit varies within a state depending on whether a store is an early opener or a late opener in the state. Adding controls for population density and controls makes little difference. For example, the difference between the 11-15 group and the 1-2 group is .63 in the regression and 3.5 - 2.9 = .6 in the raw data. These differences between early openers and late openers is highly statistically significant.

6 Bounding Density Economies

There remain three parameters to pin down, $\theta = (\tau, \omega_1, \omega_2)$, all relating to density of one form or another. The τ parameter is the coefficient on distance between a store and its distribution center in distribution costs. This parameter captures benefits of *store density*. The parameters ω_1 and ω_2 relate to how fixed cost varies with *population density*. They are the coefficients on log population density and its square in the fixed cost specification (1).

As discussed in the introduction, the cost of distance τ is intended to go beyond just the physical costs of moving goods. It is also meant to capture the idea when a store and its distribution center are far apart, it is difficult to operate the Wal-Mart model of quick response to demand shocks. It might be possible to get information about trucking costs (thought certainly not from Wal-Mart). But it would be difficult to directly measure the indirect ways that distance impedes the Wal-Mart way of doing business. Analogously, while it certainly possible to take account the higher land prices and higher wages in big cities (and I do this), it is difficult to directly measure some of the disadvantages alluded to earlier of implementing the Wal-Mart model in big cities. So rather then try to estimate these parameters through direct measurement, the approach taken here is to infer the parameters from the way Wal-Mart behaves.

6.1 The General Method

There is a large literature on bounds estimation including influential work by Manski. (See, for example, Manski (2003)). The revealed preference approach taken here generates a set of inequalities. A bounds estimation strategy is a natural way to extract information from the set of inequalities implied by choice behavior. In my implementation of this strategy, I follow Pakes, Porter, Ho, and Ishii (2006).

Recall that action a denotes a particular choice of Wal-Mart, a particular feasible solution to problem (2). Let a_0 denote the choice Wal-Mart actually made. For each policy a, let $\Pi(a)$ be the present value at date $t = 1$ of operating profits from both general merchandise and food over all stores and all time periods given policy a,

$$\Pi(a) \equiv \sum_{t=1}^{T} (\rho_t \beta)^{t-1} \left(\sum_{j \in B_t^{Wal}(a)} \pi_{jt}^g(a) + \sum_{j \in B_t^{Super}(a)} \pi_{jt}^f(a) \right). \tag{9}$$

Analogously, let $D(a)$ be the present value of all distribution miles. This is the same as the formula for $\Pi(a)$ except the distribution distance $d_{jt}^k(a)$ of store j in year t in segment k replaces the operating profit $\pi_{jt}^k(a)$. Similarly, let $F_1(a)$ be the present value of the (log) population density for each store and year and let $F_2(a)$ be the present value of the square of (log) population density. (Recall specification (1) for how fixed cost varies with population density.) Then rewriting (2), the value to Wal-Mart of choosing policy a, given a vector of density parameters $\theta = (\tau, \omega_1, \omega_2)$, is

$$v(a, \theta) = \Pi(a) - \tau D(a) - \omega_1 F_1(a) - \omega_2 F_2(a)$$

Let θ_0 be the true parameter. The policy a_0 chosen by Wal-Mart solves problem (2). Hence at θ_0,

$$v(a_0, \theta_0) \geq v(a, \theta_0), \text{ for all } a \neq a_0$$

Or

$$\Delta v(a, \theta_0) \geq 0,$$

for

$$\Delta v(a, \theta) \equiv v(a_0, \theta) - v(a, \theta). \tag{10}$$

We can decompose this as

$$\Delta v(a, \theta) = \Delta \Pi(a) - \tau \Delta D(a) - \omega_1 \Delta F_1(a) - \omega_2 \Delta F_2(a).$$

In the econometrics to follow, the error term will arise on account of measurement error. Recall that operating profit in market segment k at a particular store j in period t given some policy a can be written in the form

$$\pi_{jt}^k = (\mu - \nu_{other}) R_{jt}^k - w_{jt} \nu_{labor} R_{jt}^k - r_{jt} \nu_{land} R_{jt}^k.$$

As explained in Section 4, there is measurement error in the wage and land-rent measures. So the observed store operating profit is

$$\tilde{\pi}_{jt}^k = (\mu - \nu_{other}) R_{jt}^k - \left(w_{jt} + \varepsilon_{jt}^{wage}\right) \nu_{labor} R_{jt}^k - \left(r_{jt} + \varepsilon_{jt}^{rent}\right) \nu_{land} R_{jt}^k,$$

for measurement error ε_{jt}^{wage} and ε_{jt}^{rent}. Assume ε_{jt}^{wage} is mean zero and independent of R_{jt}^g, d_{jt}^g, R_{jt}^f, d_{jt}^f, $f_{1,jt}$ and $f_{2,jt}$. Make the analogous assumption on ε_{jt}^{rent}. Aggregate across stores and time like in (9) to get the present value of observed operating profits $\tilde{\Pi}(a)$ for each action. Substitute this into the analog of (**??**) to get the observed value $\tilde{v}(a, \theta)$ under policy a given θ. Like (**??**), let $\Delta \tilde{v}(a, \theta)$ be the observed difference in value between the chosen policy a_0 and some other policy a. It can be written as

$$\Delta \tilde{v}(a, \theta) = \Delta \Pi(a) - \tau \Delta D(a) - \omega_1 \Delta F_1(a) - \omega_2 \Delta F_2(a) + \eta_a,$$

for $\eta_a = \varepsilon_{a_0} - \varepsilon_a$, and

$$\varepsilon_a \equiv \sum_{t=1}^{T} (\rho_t \beta)^{t-1} \left(\sum_{j \in B_t^{Wal}(a)} \left(\varepsilon_{jt}^{wage} \nu_{labor} + \varepsilon_{jt}^{rent} \nu_{land}\right) R_{jt}(a) \right). \tag{11}$$

To ease the notational burden, let $y_a = \Delta \Pi(a)$, $x_a = (\Delta D(a), \Delta F_1(a), \Delta F_2(a))'$.and $\theta = (\tau, \omega_1, \omega_2)$. Then

$$\Delta \tilde{v}(a, \theta) = y_a - x_a \theta + \eta_a. \tag{12}$$

Given the assumptions made on ε_{jt}^{wage} and ε_{jt}^{rent}, the composite measurement error η_a is mean zero and independent of x_a. (It is independent of y_a as well, but that is not particularly helpful since only $y_a + \eta_a$ is observed.) Note that at the true parameter θ_0,

$$y_a - x_a\theta \geq 0, \text{ for all feasible } a.$$

Consider a set of feasible deviations defined in a manner that is unrelated to the measurement error η_a. Let there be N such sets indexed by i and let A_i denote the i-th set. Let $w_{a,i} \geq 0$ be weighting variables. Define the *basic moment* for each i by

$$m_i(\theta) = \sum_{a \in A_i} w_{a,i} \Delta \tilde{v}(a, \theta). \tag{13}$$

Next consider more complicated moments that are derived from interactions. Let x_a^k be the k-th element of x_a and suppose a lower bound $\underline{x}_a^k$ exists so that $x_a^k \geq \underline{x}_a^k$ for all a. Define

$$\underline{z}_a^k \equiv x_a^k - \underline{x}^k \geq 0.$$

Analogously, suppose an upper bound $\bar{x}^k$ exists so that $x_a^k \leq \bar{x}^k$ for all a. Define

$$\bar{z}_a^k \equiv \bar{x}^k - x_a^k \geq 0.$$

For each i define the *interaction moments* by

$$\begin{aligned} \underline{m}_i^k(\theta) &= \sum_{a \in A_i} w_{a,i} \Delta \tilde{v}(a, \theta) \underline{z}_a^k, \; k \in \{1, 2, 3\} \\ \bar{m}_i^k(\theta) &= \sum_{a \in A_i} w_{a,i} \Delta \tilde{v}(a, \theta) \bar{z}_a^k, \; k \in \{1, 2, 3\} \end{aligned} \tag{14}$$

The $\underline{z}_a^k$ and $\bar{z}_a^k$ are valid instruments because they are uncorrelated with the measurement error. If we think about $\Delta \tilde{v}(a, \theta)$ as a kind of residual, the use of the interaction moments here is analogous the familiar moment conditions for OLS, $0 = (Y - Xb)'X$.

There are N basic moments (one for each set A_i) and $3N$ interaction moments for a total of $4N$ moments. Stack the moments in a vector $M(\theta)$ and index the moments by h. At the true parameter θ_0,

$$E[M_h(\theta)] \geq 0, \text{ for all } h. \tag{15}$$

The set of θ satisfying these moment inequalities is the identified set. To estimate θ_0, find the values that satisfy the sample analog of the moment inequalities. If no such values exist, then, following Pakes et al. (2006), take the values that are closest to satisfying the inequalities. In particular, the estimate $\hat{\theta}$ solves

$$\hat{\theta} = \arg\min \sum_{h=1}^{4N} M_h^-(\theta)^2$$

for

$$M_h^-(\theta) = \min \{M_h(\theta), 0\}.$$

In estimation below, there exists a parameter region in which all $4N$ moment inequalities are satisfied. It is still necessary, however, to specify what happens when there is no solution satisfying all the inequalities, because the issue comes up while simulating the standard errors.

The focus of the estimation is bounding the distribution cost parameter τ. An upper bound $\hat{\tau}^{upper}$ is obtained by maximizing τ over the set of feasible $\theta = (\tau, \omega_1, \omega_2)$. The feasible θ satisfy the sample analogs of (15) as well as restrictions $\omega_1 \geq 0$, $\omega_2 \leq 0$ (so the fixed cost (1) is weakly concave in log population density. A lower bound $\hat{\tau}_{lower}$ is obtained similarly. Linear programming techniques are employed. Pakes and el. (2006) provide a method of simulating confidence intervals for $\hat{\tau}^{upper}$ and $\hat{\tau}_{lower}$. They describe two statistics that bracket the distribution. As a first step to having something to say about confidence intervals, I simulate these confidence intervals exactly as they propose.

There are approximately 5000 store openings. (This figure treats the food business in a supercenter as a different operation with its own opening date.). By interchanging the openings of pairs of stores, there are millions of possible deviations based on these 5,000 openings. In the theory underlying Pakes et al(2006), the sample size equals the number of deviations. But in my case there are only 5000 store openings and the measurement errors underlying the millions of deviations are different combinations of differences across these 5000 stores. This leads me to consider a modification of Pakes et al (2006) procedure in which for each simulation run, random draws are obtained at the level of the store and the draws for the various deviations are obtained from differencing the store-level draws.

7 Specifics of Method

Like Bajari and Fox (2005) and Fox (2005), I restrict attention to pairwise resequencing. I consider only deviations a in which the opening dates of pairs of stores is reordered. For example, store #1 actually opened 1962 and #2 opened 1964. Three classes of subsets are constructed.

1. *Store-Density Decreasing.* Start with the set of stores opened 10 or more years after the first store in their state. For each such store j, find the set of stores, indexed by j', such that store j' opens three or more years after store j in a different state. Furthermore, require that store j' be opened within four years of the first store in j''s state. This is set A_1.

2. *Store-Density Increasing.* Start with the list of stores opened within five years of the first store in the state. For each such store j, find the set of stores, indexed by j', such that store j' opens three or more years after store j in a different state. Furthermore, require that store j' be opened 10 or more years after the first store in j''s state. Next, require that the first store in j''s state is before the first store in j's state. Flipping the opening order, this is set A_2

3. *Population-Density Increasing (Decreasing).* Define store locations by population density groupings. Let class 1 be locations with less than 15 (thousand) in population within five miles, class 2 be 15 to 40, class 3 be 40 to 100, and class 4 over 100. Take store locations in the same state opening at different dates where one location is in population density class and the other is in a different class. Flip the order of store openings. The various combinations result in 6 different sets of deviations.

The purpose of set 1 is to provide information for a lower bound on $\hat{\tau}_{lower}$. In the deviation, instead of adding yet another store in a state where Wal-Mart has been for over 10 years, store that would have been one of the early stores in a another state is opened even sooner. The alternative location is one which has not yet been hit with diminishing returns. Set 2 goes the other direction and provides the upper bound on $\hat{\tau}^{upper}$. The third category defines pairwise perturbations based on population density that are intended to provide information about the parameters ω_1 and ω_2 that govern how the fixed cost varies with population density.

7.1 Estimates

Consider the set of *store-density decreasing* deviations defined above by A_1. In each of these, Wal-Mart is delaying the opening of a store in its existing network and opening sooner a store farther out from its network. Given the parameters defining the deviation (e.g. start with the set of stores opened 10 years after the first store in the state, etc.), there are 239,698 such deviations Wal-Mart could have considered that reorder store opening dates and 5,110 that reorder supercenter conversion dates. Calculating the sales model is time consuming so I take a large random sample of the $a \in A_1$ (approximately 83,000) of this set to estimate the means. Table 10 reports summary statistics for the sample of A_1 deviations as well as the 7 other deviation sets. The weight variable scales changes in value from the deviation into present value terms as of the year of the earliest store being switched. This puts on equal footing deviations taking place at different time periods.

According to Table 10, the average change in discounted operating profit across the type 1 deviations (store-density decreasing) was -\$3.8 million. The negative value implies that by going with the chosen policy a_0 rather than a deviation $a \in A_1$, on average Wal-Mart was losing money. So random deviations to disperse stores raises the present value of profits by \$3.8 million. (Recall that these is measurement error on wages and rents, but this should average out.) These deviations come at a cost, since mean $\Delta D = -1740$. This means that annual store distribution miles (discounted by β) increase on average 1,740 miles.

Next consider type 2 deviations which increase store density. By following the actual policy a_0 rather than deviate this way, Wal-Mart enjoyed $\Delta\Pi = 4.8$ more in operating profits than on average from these deviations. But it was also burdened by $\Delta D = 1,205$ in store-year distribution miles.

For now, ignore fixed cost from population density by assuming $\omega_1 = 0$ and $\omega_2 = 0$. We can then use the first two moment inequalities to bound τ. Since Wal-Mart chose a_0, the expected value of the first moment inequality must be nonnegative at the true τ,

$$\begin{aligned} E[M_1] &= \Delta\Pi_1 - \tau\Delta D_1 \\ &= -3.8 + \tau 1740 \\ &\geq 0 \end{aligned}$$

This implies that

$$\tau \geq \hat{\tau}_{lower} = \frac{3.8}{1740} = .0218(\ \$\text{million}) = \$2,180. \tag{16}$$

This says that the cost of one store distribution mile is at least \$2,180. Analogously, we can use type 2 deviations to derive an upper bound of

$$\tau \leq \hat{\tau}^{upper} = \frac{4.8}{1201} = .0400(\$\text{million}) = \$4,000.$$

These findings are intriguing but incomplete as they don't take into account the population density component of first costs. Note that by doing the actual policy a_0 rather than the type 1 or type 2 deviations, Wal-Mart tended to stay out of higher population density locations longer (ΔF_1, the difference in present value log population density from doing a_0, is negative). If Wal-Mart is putting weight ω_1 on this, then that can explain part of the reason why Wal-Mart was willing to give up operating profits by doing a_0 rather then deviate in the type 1 way. Inequality (1) gives all of the credit to τ in getting moment equality one to hold. If we increase ω_1, we can lower τ and the moment inequality still holds. Analogously, increasing ω_1 puts slack into $E[M_2]$ allows us to raise $\hat{\tau}^{upper}$. Now if we free to pick any ω_1 and ω_2 then moments 1 and 2 would put no restriction on τ. This is where moments 3 through 8 come in. With these deviations, the stores being flipped are in the same state so the impact on distribution miles is minimal, as can be seen in the ΔD column in the table. These deviations do change population density and so put discipline on the choice of parameters ω_1 and ω_2.

Table 11 presents the results of moment inequality estimation when the full set of constraints is imposed. (There are eight basic moments plus 3 interaction moments for each for a total of $32 = 8 \times 4$ moments.) The linear programming of minimizing or maximizing τ subject to the constraints is solved. The exercise is conducted with all store openings and then a breakdown by whether early opening store in the deviation was opened after 1990 or before 1990. The period after 1990 is particularly interesting that is when supercenters began to open.

For each estimation, four sets of confidence intervals are reported. Set 1 corresponds to what Pakes et al (2006) call the *inner* approximation of distribution. It is constructed by simulating draws from data (y_a, x_a) from which the moments are defined and then running the full linear programming estimation with the simulated data over the 32 constraints. Set 2 is the *outer* approximation. It is derived using the parameter vector $\hat{\theta}_{lower} = (\hat{\tau}_{lower}, \hat{\omega}_{1,lower}, \hat{\omega}_{2,lower})$ for which the minimum value of τ is obtained and using this to simulate data. A set of estimates of τ is derived but rather than solve the full linear program only the three binding constraints from the original estimate are imposed (and they are required to be binding). Sets 3 and 4 are incorporated as a way to take into account that the draws across the different deviations are not i.i.d. because there are only 5000 store openings we are taking differences from. The strategy is to use a version of Pakes et al's (2006) outer procedure. But rather than draw the right-hand side variables randomly from a limiting normal distribution, the strategy uses a bootstrap procedure. Set 3 is exactly the same as Set 2, except for the bootstrap. Table 11 shows that for the most part, use of the bootstrap rather than simulated draws does not make a big difference between set 2 and set 3. Finally, set 4 is just like set 3, except that in simulation it takes into account that draws are not i.i.d.. Rather, it takes draws for the 5000 store openings and then takes differences. So for example, there are 5000 draws of data rather than 83,000 draws of data on moment 1. Taking into account there is less data because things are not i.i.d. does increase the variance, as expected. (That is, moving from set 3 to set 4 widens the confidence intervals.) Nonetheless, it is clear that even taking this into account, there is a lot of information here and it is possible to place meaningful bounds on the distribution of the τ with a significant degree of confidence.

More to analysis to come!

- Intepretation of the estimate of τ
- A comparison to an estimate of τ obtained by changing the date of opening of a distribution center. (A similar order of magnitude.)

References

References

[1] Andrews, Donald W.K., Steven Berry, and Panle Jia, (2004) "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Yale University working paper.

[2] Bajari, Pat, Benkard, L., and J. Levin, "Estimating Dynamic Models of Imperfect Competition, " Stanford University Working Paper, 2004.

[3] Bajari, Patrick, and Jeremy Fox (2005), "Complementarities and Collusion in an FCC Spectrum Auction," University of Chicago Working Paper

[4] Basker, Emek,"Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion" *forthcoming Review of Economics and Statistics.*

[5] Bresnahan and Reiss, "Entry and Competition in Concentrated Markets" Journal of Political Economy v99, n5 (Oct 1991):977-1009.

[6] Ellickson, Paul E., "Does Sutton Apply to Supermarkets?" Forthcoming in the Rand Journal of Economics. Spring, 2007

[7] Fox, Jeremy, (2005), "Semi and Nonparametric Estimation of Multinomial Discrete Choice Models Using a Subset of Choices," University of Chicago Working Paper

[8] Ghemawat, Pankaj, Ken A. Mark, Stephen P. Bradley,(2004) "Wal-Mart Stores in 2003," Harvard Business School Case Study 9-704-430.

[9] Hausman, Jerry and Ephraim Leiptag, (2005) "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart," MIT working paper.

[10] Holmes, Thomas J., "Barcodes lead to Frequent Deliveries and Superstores," Rand Journal of Economics Vol. 32, No. 4, Winter 2001, pp 708-725.

[11] Holmes, Thomas J., (2005) "The Location of Sales Offices and the Attraction of Cities," , *Journal of Political Economy*

[12] Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry" Yale University Manuscript, 2005

[13] Ledyard, Margert (2004), "Smaller Schools or Longer Bus Rides? Returns to Scale and School Choice " working paper.

[14] Manski, C., (1989), "Anatomy of the Selection Problem," *Journal of Human Resources* 24, 343-60.

[15] Neumark, David, Junfu Zhang, and Stephen M. Ciccarella, Jr. "The Effects of Wal-Mart on Local Labor Markets" NBER Working Paper No. W11782 November 2005

[16] Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii, "Moment Inequalities and Their Application," Harvard working paper, April, 2005.

[17] Roberts, Mark J. "Economies of Density and Size in the Production and Delivery of Electric Power Mark, *Land Economics*, Vol. 62, No. 4. (Nov., 1986), pp. 378-387.

[18] Stone, Kenneth E., "Impact of Wal-Mart Stores on Iowa Communities: 1983-93," Economic Development Review, Spring 1995, p. 60-69.

[19] Toivanen, Otto and Michael Waterson (2005), "Market Structure and Entry: Where's the Beef?" *Rand Journal of Economics,* Vol 36, No. 3,

[20] Walton, Sam (1992) *Made in America: My Story* New York: Doubleday

[21] Vance, Sandra S. and Roy Scott, *Wal-Mart: A History of Sam Walton's Retail Phenomenon*, Simon and Schuster Macmillan: New York, 1994.

Table 1
Summary Statistics of TradeDimensions Wal-Mart Data
(End of 2005, Excludes Alaska and Hawaii)

Store Type	Variable	N	Mean	Std. Dev	Min	Max
All	Sales ($millions/year)	3,176	70.5	30.0	9.1	166.4
Regular Store	Sales ($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 Store	Employment	1,196	123.5	40.1	57.0	410.0
SuperCenter	Employment	1,980	333.8	91.5	31.0	812.0

Table 2
Distribution of Wal-Mart Facility Opens by Decade and Opening Type

	Wal-Marts		Supercenters		Regional Distribution Centers		Food Distribution Centers	
Decade Open	Opened this decade	Cumulative	Opened this decade	Cumulative	Opened this decade	Cumulative	Opened this decade	Cumulative
1962-1969	15	15	0	0	1	1	0	0
1970-1979	243	258	0	0	1	2	0	0
1980-1989	1,082	1,340	4	4	8	10	0	0
1990-1999	1,130	2,470	679	683	18	28	9	9
2000-2005	706	3,176	1,297	1,980	14	42	25	34

Table 3
Summary Statistics for Census Block Groups

	1980	1990	2000
N	269,738	222,764	206,960
Mean population (1,000)	0.83	1.11	1.35
Mean Density (1,000 in 5 mile radius)	165.3	198.44	219.48
Mean Per Capita Income (Thousands of 2000 dollars)	14.73	18.56	21.27
Share Old (65 and up)	0.12	0.14	0.13
Share Young (21 and below)	0.35	0.31	0.31
Share Black	0.10	0.13	0.13

Table 4
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	1.938
		(.056)	(.043)
λ^f	Food spending per person (annual in $1,000)	1.649	1.912
		(.061)	(.050)
δ_0	Distance disutility (constant term)	.642	.703
		(.036)	(.039)
δ_1	Distance disutility (coefficient on population density)	-.046	-.056
		(.007)	(.008)
ω	Outside Alternative valuation parameters		
	constant	-8.271	-7.834
		(.508)	(.530)
	ln(mbar)	1.968	1.861
		(.138)	(.144)
	ln(mbar)^2	-.070	-.059
		(.012)	(.013)
	Per Capita Income	.015	.013
		(.003)	(.003)
	Share of block group black	0.341	.297
		(.082)	(.076)
	Share of block group young	1.105	1.132
		(.464)	(.440)
	Share of block group old	.563	.465
		(.380)	(.359)
γ	Store-specific parameters		
	store age 2+ dummy	.183	.207
		(.024)	(.023)
σ^2	measurement error	.065	.065
		(.002)	(.002)
N		3176	3176
SSE		205.117	206.845
R^2		.755	0.753
ln (L)		-155.749	-169.072

Table 5
Cannibalization Rates
From Annual Reports and in Model

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*

*Cannibalization Rate is imposed to equal 1.00 in 2005.

Table 6
Comparative Statics with Demand Model
(Uses Constrained Model)

Distance	Population Density (thousands of people within a 5 mile radius)						
(miles)	1	5	10	20	50	100	250
0	.999	.989	.966	.906	.717	.496	.236
1	.999	.979	.941	.849	.610	.387	.172
2	.997	.962	.899	.767	.490	.288	.123
3	.995	.933	.834	.659	.372	.206	.086
4	.989	.883	.739	.531	.268	.142	.060
5	.978	.803	.615	.398	.184	.096	.041
10	.570	.160	.083	.044	.020	.011	.006

Table 7
Distribution of Variable Input Costs
(Percentiles of Distribution are Weighted by Sales Revenue)

Estimated 2005 Labor Costs

Quartile	Store Location	Annual Payroll per Worker ($)	Wages as Percent 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 Land-Value/Sales Ratios (Expressed as a Percent)

Quartile	Store Location	Index of Residential Property Value Per Acre ($)	Land-Value/Sales Ratio
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

Table 8
(All evaluated at 2005 Demand Equivalents)

Part A: General Merchandise (New Wal-Marts including supercenters)

	N	Incremental Sales	Incremental Operating Profit	Incremental Distribution Center Distance (miles)	Stand-alone Sales	Stand-alone Operating Profit
All	3,176	36.3	3.1	168.9	41.4	3.6
By State's Wal-Mart Age at Opening						
1-2	288	38.0	3.5	343.3	38.7	3.6
3-5	614	39.5	3.5	202.0	41.5	3.7
6-10	939	37.6	3.3	160.7	40.9	3.6
11-15	642	36.1	2.9	142.1	42.2	3.4
16-20	383	32.9	2.8	113.7	41.2	3.5
21 and above	310	29.5	2.4	90.2	44.4	3.6

Part B: Food (New supercenters)

	N	Incremental Sales	Incremental Operating Profit	Incremental Distribution Center Distance	Stand-alone Sales	Stand-alone Operating Profit
All	1,980	40.2	3.6	137.0	44.8	4.0
By State's Wal-Mart Age at Opening						
1-2	202	42.4	3.9	252.9	3.9	4.0
3-5	484	42.7	4.0	171.2	4.1	4.2
6-10	775	41.0	3.6	113.5	4.0	4.0
11-15	452	36.7	3.2	95.3	3.9	3.9
16-20	67	30.1	2.8	94.0	3.5	3.5

Table 9
Incremental Operating Profit Regression
2005 Demand Equivalents
Includes State Fixed Effects

	General Merchandise	Food
State's Wal-Mart Age at Opening		
3-5	-0.04	-0.10
	(.05)	(.07)
6-10	-0.30	-0.60
	(.05)	(.07)
11-15	-0.63	-1.10
	(.05)	(.08)
16-20	-0.76	-1.35
	(.06)	(.12)
21 plus	-1.32	
	(.06)	
log population density	5.78	6.39
	(.18)	(.31)
(log population density)^2	-0.26	-0.28
	(.01)	(.01)
R^2	.51	.50
N	3176	1980

Table 10

Summary Statistics by Type of Deviation

Deviation Type	Description	Sample Number	Mean Values $\Delta\Pi$	ΔD	ΔF_1	ΔF_2
1	Store-Density Decreasing	83,625	-3.8	-1,740.3	-2.1	-13.1
2	Store-Density Increasing	109,452	4.8	1,201.5	-4.8	-37.7
	Population-Density Changing					
3	Class 4 to Class 3	5,579	1.6	-9.9	3.6	34.8
4	Class 3 to Class 2	8,226	4.4	25.0	3.5	26.6
5	Class 2 to Class 1	12,176	5.0	-64.3	3.2	17.4
6	Class 1 to Class 2	10,182	-2.2	-53.3	-3.5	-19.8
7	Class 2 to Class 3	12,110	1.0	-89.7	-4.0	-29.9
8	Class 3 to Class 4	13,164	2.7	-28.1	-4.9	-47.3

Table 11

Estimated Bounds on Distribution Cost τ
Units are Dollars per Mile Year
(95 Percent Confidence Intervals in Parentheses Constructed Four Ways)

Sample of Perturbations Used for Estimate	Lower Bound	Upper Bound
All Openings	724 #1(700,4810) #2(530,910) #3(700,740) #4(550,910)	6712 #1(2940,6990) #2(5560,8880) #3(6620,6800) #4(6120,7290)
Openings 1990 and beyond	1780 #1(1820,4170) #2(1730,1830) #3(1640,1910) #4(940,2630)	5190 #1(1950,5310) #2(4370,7930) #3(5140,5250) #4(4850,5520)
Openings before 1990	590 #1(570,4660) #2(420,760) #3(560,620) #4(330,850)	8330 #1(6800,9180) #2(7000,13660) #3(8180,8510) #4(7340,9350)

Figure 1
Sales per Employee As a Function of Square Footage of Store