

The research program of the Center for Economic Studies (CES) produces a wide range of theoretical and empirical economic analyses that serve to improve the statistical programs of the U.S. Bureau of the Census. Many of these analyses take the form of CES research papers. The papers are intended to make the results of CES research available to economists and other interested parties in order to encourage discussion and obtain suggestions for revision before publication. The papers are unofficial and have not undergone the review accorded official Census Bureau publications. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the U.S. Bureau of the Census. Republication in whole or part must be cleared with the authors.

**SUPERSIZE IT:**

**THE GROWTH OF RETAIL CHAINS AND THE RISE OF THE “BIG BOX”  
RETAIL FORMAT**

by

**Emek Basker \***  
**University of Missouri**

**Shawn Klimek \***  
**U.S. Bureau of the Census**

and

**Pham Hoang Van \***  
**Baylor University**

**CES 08-23    August, 2008**

All papers are screened to ensure that they do not disclose confidential information. Persons who wish to obtain a copy of the paper, submit comments about the paper, or obtain general information about the series should contact Sang V. Nguyen, Editor, [Discussion Papers](#), Center for Economic Studies, Bureau of the Census, 4600 Silver Hill Road, 2K132F, Washington, DC 20233, (301-763-1882) or INTERNET address [sang.v.nguyen@census.gov](mailto:sang.v.nguyen@census.gov).

## Abstract

We offer a theory for the complementarity between the size of a retail chain and the scope of its business to explain the growth of general-merchandise firms and the expansion of the “superstore” format. The complementarity results from an interaction of the retailer’s economies of scale and consumer gains from “one-stop shopping.” We find support for our model in micro data from the Census of Retail Trade for 1977–2002. Retail chains with more stores carry more distinct product lines and as retail chains grow they add both stores and product lines. On average, we find that a chain adds one product line, such as shoes, computers, or jewelry, to an existing store with every new store it opens. For the average large chain, adding a new product line throughout the chain is correlated with adding 400 new stores, competing in over 8,000 new markets and increasing its competitive pressure in more than 10,000 additional markets.

JEL Codes: L11, L25, L81

Keywords: Retail, Chain, Big Box, Superstore, Economies of Scale, General Merchandise, One Stop Shopping

\* Comments welcome to: [emek@missouri.edu](mailto:emek@missouri.edu), [shawn.d.klimek@census.gov](mailto:shawn.d.klimek@census.gov), or [vanpham@baylor.edu](mailto:vanpham@baylor.edu). The opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Otherwise, this work has not undergone the review accorded official Census Bureau publications. We thank Saku Aura, Roger Betancourt, Jonathan Beck, Tom Davidoff, Lucia Foster, David Gautschi, Tom Holmes, Ron Jarmin, Julia Lane, David Mandy, and seminar participants at Berkeley (Haas), Census Bureau, Missouri, the 2006 “Advances in the Empirical Analysis of Retailing” workshop at WZ-Berlin, the 2007 AEA (Chicago), the 2007 IIOC (Savannah), and the 2008 RPI Mini-Conference on Marketing and Innovation for comments. Basker also thanks the University of Missouri Economic and Policy Analysis Research Center (EPARC) and the University of Missouri Research Board for financial support.

# 1 Introduction

A major source of growth in the retail sector has been “big box” retail chains: general merchandisers selling everything from toothpaste to baby strollers to tires and toasters. Between 1977 and 2002 the share of all retail sales accounted for by general-merchandise retailers grew from about 10% to almost 15%. The top four general-merchandise firms now account for almost 10% of all retail sales, more than double their share twenty five years ago. These chains’ growth has been fueled by expansion on two margins: they operate more stores than ever before, and they sell a larger breadth of assortment in each store. Wal-Mart, for example, introduced its “Supercenter” store format, which includes a full line of groceries, in 1988, and by 2002 was the largest grocer in the U.S. (Basker, 2007); Kmart and Target have introduced similar formats called SuperTarget stores and Kmart Super Centers, respectively. These large chains’ growing dominance has contributed to productivity growth in the retail sector and increased the competitive pressures faced by many incumbent businesses.

In this paper, we argue that the sharp growth of the general-merchandise subsector is due to a complementarity between retail chains’ scale (number of stores) and scope (number of products). We offer a model in which a chain’s choice of product scope depends positively on the number of stores it operates and vice versa. The complementarity between the number of stores and the number of products sold at each store results from an interaction of economies of scale on the cost side with economies of scale on the demand side. On the cost side, as a chain adds stores — increasing sales volume — economies of scale imply a falling cost per unit sold. The lower marginal cost induces the chain to add product offerings, which draw in more customers due to a “one-stop shopping” effect (economies of scale in demand) by allowing consumers to save on transportation costs. The combination of more products and more customers at each store increases store profit, prompting the retail chain to add even more stores. We show that when consumers’ transportation cost is high enough the retailer’s scale and scope are strategic complements and move together in response to changes in transportation cost, the wholesale prices facing Mom and Pop stores, and technological

innovations.

We use newly-developed micro data from the Census of Retail Trade to test the predictions of the model. The empirical analysis is based on firm-level records for all general-merchandise retailers from the Census of Retail Trade spanning six censuses, from 1977 to 2002. We calculate, for each retailer in each census year, both its scale (measured as the number of stores it operates) and its scope (the number of product lines it sells). We use a difference-in-difference specification to document a strong and persistent relationship between the size of a retail chain and its scope. As retail chains grow, they add both stores and product lines; we find that when a general merchandiser opens a new store, one store in the chain adds a new product line, such as shoes, computers, or jewelry.

For large chains this interaction between scale and scope has far-reaching effects on competition. In our data, the average large chain (defined as having at least 100 stores) operates more than 500 stores and sells approximately 20 product lines. When the average large chain adds another product line to its slate, whether due to innovations in its inventory technology or to changes in consumer demand parameters, economies from the scale-scope interaction also induce it to open another 400 stores. At 20 product lines per store, this is an expansion into 8,000 new markets, or an increase of 55% in the number of markets served by the chain. In addition, the one-stop shopping effect makes the chain more competitive in the 20 product lines already sold in the existing 500 stores, indirectly affecting more than 10,000 additional markets.

To our knowledge, this paper is the first to consider the interaction between scale and scope in the retail sector. Existing work on chains mostly focuses on the spatial aspects of a chain's location decisions (see, for example, [Jia, forthcoming](#); [Holmes, 2008](#)) and abstracts from the chain's choice of product selection. Work on large or multi-product retail stores, in contrast, tends to treat stores as independent entities or ignores the distinction between chains and store altogether. [Bagwell, Ramey, and Spulber \(1997\)](#), for example, analyze the investment decisions of retail firms while abstracting from both the number of locations they

operate and the range of products they sell. [Holmes \(2001\)](#) explicitly models a retailer’s endogenous store size as a function of inventory-management technology such as bar codes or radio frequency identification techniques, but treats the store as an independent entity, with no role for a retail chain. Similarly, [Bliss \(1988\)](#) models competition between “specialist” shops and a “general” shop without reference to chains. An important exception is [Betancourt \(2004, pp. 173-180\)](#), who explicitly models a chain’s choice of scope and notes the symmetries between scope and scale.

The rest of the paper is organized as follows. Section 2 provides an overview of the general-merchandise subsector and the changes it has undergone in the last few decades, as it compares to the retail sector as a whole. In Section 3 we offer a model of chain retailing and competition and derive conditions for complementarity of a dominant retailer’s store size (scope) and its chain size (scale). We describe the Census micro data in Section 4 and present our empirical findings, which confirm the relationship between chain scale and scope, in Section 5. Section 6 concludes.

## 2 Background and Motivation

We use the following terminology in our discussions. A **retailer** is a firm that operates one or more retail establishments, or **stores**. A retailer operating a single store is a **single-store retailer**. A retailer operating multiple stores is a **retail chain**. We further distinguish between a **large retail chain**, which operates 100 or more stores, and a **small retail chain**, which operates 2–99 stores.

Our focus in this paper is on the general-merchandise retail subsector, a classification that includes department stores, discount department stores, dollar stores, general stores, variety stores, and trading posts: stores selling a variety of products.<sup>1</sup> We refer to general

---

<sup>1</sup>The North American Industrial Classification System (NAICS), currently used by the Census, defines “General Merchandise Stores” as stores that sell “new general merchandise [to retail consumers] from fixed point-of-sale locations. Establishments in this subsector are unique in that they have the equipment and

merchandisers, especially ones selling many products, as **big-box stores** or **superstores**. We call other retailers **specialists**. A single-store specialist is a “**Mom and Pop**” retailer. A general merchandiser carries multiple **product lines** (which we also refer to as products or lines). A **market** is location and product pair; a single-store retailer selling five products competes in five markets, while a chain of 100 stores selling five products competes in 500 markets.

Table 1 shows the number of general-merchandise firms in the Census of Retail Trade (CRT) each census year since 1977, by retailer size.<sup>2</sup> Over the course of the quarter century from 1977 to 2002, the number of single-store general merchandisers has fallen by nearly 40% and the number of small chains by over 75% while the number of large chains has remained virtually unchanged. This consolidation is not a symptom of a shrinking subsector, however: sales in real terms have increased by more than 50%. Chains’ dominance in general merchandising is far greater than in the rest of the retail sector. Table 2 provides a snapshot of the differences between the retail sector as a whole and the general-merchandise subsector from the 2002 CRT. General-merchandise retailers constitute only 1.3% of all retail firms, but account for 3.7% of retail stores and 14.6% of retail sales. In 2002, while 38% of retail stores belonged to chains (most of them small), 78% of general-merchandise stores belonged to chains and 72% to chains with at least 100 stores.

The differences between the general-merchandise subsector and the rest of the retail sector are not new, but they have increased over time. Figure 1(a) documents the rising share of retail sales accounted for by retail chains.<sup>3</sup> Until the late 1970s, more than half of

---

staff capable of retailing a large variety of goods from a single location” ([Office of Management and Budget, 1998](#)). Before the classification change that took place in 1997, the Census used the Standard Industrial Classification (SIC) system, which defined the subsector as consisting of “stores which sell a number of lines of merchandise, such as dry goods, apparel and accessories, furniture and home furnishings, small wares, hardware, and food” ([Office of Management and Budget, 1987](#)).

<sup>2</sup>The CRT is part of the Economic Census, which is conducted in years ending in “2” and “7” for all establishments with paid employees.

<sup>3</sup>Figures 1(a), 1(b), and 2 were created from published CRT tables and not directly from the micro data we use later on in this paper.

all consumer dollars were spent at single-store retailers; today, more than 60% of consumer dollars are spent at chain stores, double the share of 1954. The revenue share of large chains has grown even faster than the chains' share as a whole; large chains' share of retail sales has more than tripled since the 1950s.<sup>4</sup> Chains are even more dominant in general merchandising; in terms of chain dominance, the retail sector today looks like the general-merchandise subsector more than half a century ago. Figure 1(b) shows the share of dollars spent at chains and at large chains in the general-merchandise subsector. Chains now account for virtually all general-merchandise sales: almost 99% of all general-merchandise dollars are spent in chains, and 96% in chains with at least 100 stores (up from 34% fifty years ago).

Within the general-merchandise subsector, larger chains carry more product lines. There are a total of roughly 40 broadly-defined product lines, including groceries, men's apparel, furniture, and jewelry. (These data are described in detail in Section 4. The full list of broad lines appears in Table B-1.) The raw correlation coefficient between chain size and number of products it carried ranges from 0.09–0.18 depending on whether we do the analysis in levels or logs, and are always statistically significant.<sup>5</sup> To get a concrete sense of these correlations, we calculate the average number of products lines carried by single-store retailers, small chains, and large chains. These are, respectively, twelve, fifteen, and twenty one.

As general-merchandise chains have grown, so has the number of product lines carried by their stores. Figure 2 shows the fraction of general-merchandise stores carrying selected lines in 1977 and 2002. The share of general-merchandise stores carrying groceries has increased only modestly, from 83% to 88%. The share carrying furniture, in contrast, has increased by 43%, or 23 percentage points (from 42% to 65%), between 1977 and 2002. Changes in Census forms and in the definition of product lines cause variation in the observed probability that a

---

<sup>4</sup>Jarmin, Klimek, and Miranda (forthcoming) use data from the Census Bureau's Longitudinal Business Database to document the growth of retail chains using employment per store as a measure of store size.

<sup>5</sup>The correlation coefficient is twice as large in logs as in levels.

store carries any given product, so in our empirical analysis below we create longitudinally-consistent product lines and control for changes in forms over time.

Increased competitive pressures due to big chains' expansion in both scale and scope ultimately underlie the public's seeming "love-hate" relationship with big-box chain stores. On the one hand, consumers flock to superstores to save on transportation and other transaction costs inherent in shopping in multiple stores (see, e.g., [Bonné, 2004](#); [Gogoi, 2008](#)). Wal-Mart, for example, touts the convenience of one-stop shopping in its annual reports: "That's why customers choose our Supercenters" ([Wal-Mart Stores, Inc., 1994](#), p. 7). Consumers agree: answering an open-ended question soliciting the "best thing about Wal-Mart," 50 percent of respondents of a Pew Research Center poll named low prices, 22 percent named broad selection/variety, and 13 percent gave answers related to location, hours, and other convenience factors ([Pew, 2005](#)). In the language of [Betancourt and Gautschi \(1990, 1992\)](#), one-stop shopping is a common distribution service, that is, one that affects all the goods a retailer sells.

But, at the same time, communities often resist entry by big-box chains. In recent years several states and municipalities, including the state of Maryland and the city of Chicago, have passed or attempted to pass legislation to increase the operating costs of large chains, and/or superstores, in an effort to limit their expansion. Maryland's "Fair Share Health Care Act," which would have required Wal-Mart to spend at least 8 percent of its total wage bill on health insurance, was to take effect in 2007, but was overturned by a judge in July 2006. Several other states, including New Jersey, Tennessee, and Wisconsin, have considered similar legislation. (See [Basker, 2007](#), for more details.) This aversion is not new: as far back as the 1920s, retail chains have been accused of "paying low wages, not contributing to their communities, taking money out of communities, paying fewer taxes than local merchants, and turning America into 'a nation of clerks'" ([Ross, 1984](#), p. 247).

The complementarity between scale and scope has contributed to a widespread confusion between store size and chain size. Thus, some of the discourse on big stores involves not their

size but their ownership structure. An MSNBC article entitled “Big store or small? Why consumers care,” for example, concluded that “the big-small divide rests on a basic trust equation”: consumers trust owners of small stores because they tend to be locals (Bonné, 2004). In many cases, since scale and scope are complementary, communities restricting entry use one as a proxy for the other.

The next section presents a model of a chain retailer in competition with a large number of “Mom and Pop” retailers, motivated by the above stylized facts. In equilibrium, the chain retailer’s choice of the number of products to carry and the number of stores to operate turn out to be strategic complements, and any change in parameter values — for example cost or technology parameters — leads the chain to increase (or decrease) its scale and scope together.

## 3 Model

### 3.1 Environment

There are  $N$  product categories, or lines, in the economy, and a large set of ex-ante identical locations.<sup>6</sup> There is a unit measure of consumers, each of whom has unit demand for every product line as long as the cost of procuring the item does not exceed a reservation price.

There are two types of stores: “Mom and Pop” specialist stores, each of which sells one line (such as apparel, toys, or hardware) in a single location; and a chain superstore selling multiple lines at multiple locations. Every trip to a store entails a transportation cost, which includes both the time cost and the physical transportation cost required to go to one more store. The superstore provides an opportunity for one-stop shopping by combining several shopping trips into one.<sup>7</sup> In this paper we interpret the “transportation cost” very broadly

---

<sup>6</sup>We assume that these locations do not overlap. More on this below.

<sup>7</sup>The idea that consumers incur a transportation cost to shop goes back to Hotelling (1929). Betancourt and Gautschi (1988, 1990) are among the first to rigorously model the implications of transportation costs and other distribution services for the household’s production function, but they consider only a one-store

to include any transaction cost, such as standing in line and digging for change, that needs to be repeated at every stop.<sup>8</sup>

In addition to a transportation cost, a consumer who shops at the superstore incurs an “aversion” cost representing her distaste for the “big-box” shopping environment. Consumers vary in their aversion to big-box shopping; the cost,  $\sigma$ , is distributed according to the exponential probability density function

$$f_\sigma(\sigma) = \exp(-\sigma)$$

with the rate parameter set to 1 for simplicity.<sup>9</sup>

A Mom and Pop store selling product  $i$  has constant marginal cost  $c_m^i$ . The chain’s cost function is

$$C_s = \frac{d(n, k)}{\delta} + \sum_{i=1}^n w(kx_s^i)$$

where  $\frac{d(n, k)}{\delta}$  is the home-office cost of coordination, management, logistics and distribution for a chain selling  $n$  product lines in  $k$  stores, with a technology parameterized by  $\delta$ ; and  $w(\cdot)$  is the variable cost of selling a given volume of a single line. The variable cost  $w(\cdot)$  is positive, increasing, and concave. To ensure that the superstore has lower marginal cost than the Mom and Pop stores, we assume that  $w'(0) < c_m^i$  for all  $i$ . The coordination cost  $d(\cdot, \cdot)$ , which corresponds closely to what [Bliss \(1988\)](#) calls “overhead cost,” is increasing and convex in both  $n$  and  $k$ , at least in some regions, otherwise we would get a corner solution

---

retailer. We complement that analysis by endogenizing the size of the chain along with the retailer’s product selection (a form of distribution services considered as an “implicit good” by [Betancourt and Gautschi](#)) and prices.

<sup>8</sup>Seen this way, malls are an imperfect substitute for superstores. Consumers drive to one location and park there but engage in many separate transactions.

<sup>9</sup>The exponential distribution function has support on  $[0, \infty)$ , which ensures positive demand at Mom and Pop stores even if they compete with the superstore and regardless of the price differential between these stores. We also show below that this density function guarantees a unique solution for the superstore’s profit, avoiding problems of multiple equilibria. The choice of 1 for the rate parameter does not affect any qualitative results.

in which  $n$  and/or  $k$  take on their largest possible value, as noted by [Betancourt \(2004\)](#).<sup>10</sup> In general, it is not separable in  $n$  and  $k$ . We assume that  $\frac{\partial^2 d(n,k)}{\partial k \partial n}$  exists and is continuous. This formulation implicitly assumes that all lines are symmetric with respect to their fixed costs, that all stores are symmetric, and that the increase in the cost of adding a line as the number of stores increases is the same as the increase in the cost of adding a store as the number of lines increases. We also assume that  $\frac{\partial^2 d(n,k)}{\partial k \partial n} \leq R$ , where  $R$  is a positive constant. The parameter  $\delta$  indexes the firm's technology; the larger the value of  $\delta$ , the more advanced the firm's technology with respect to inventory management, logistics, distribution, and the like.

We analyze the superstore's optimal choice of its scale (number of stores) and scope (number of product lines) using backward-induction in a three-stage process. In the first stage, the superstore selects the number of stores it operates and the number of lines it carries. In the second stage, the superstore and all Mom and Pop stores simultaneously set prices. In the third stage, consumers in each location observe their shopping choices and all prices and make their purchases. We use a Nash solution for the simultaneous price game and monotone methods to demonstrate that the chain's scope and scale are complements. The timing is summarized in Figure 3.

## 3.2 Consumers

Consumers take the presence of a superstore, the lines it carries, and all stores' prices as given when they decide where to shop. The consumer's problem is to minimize the total cost (inclusive of transportation and aversion costs) of buying all  $N$  goods. Let the price of item  $i$  at the Mom and Pop store and at the superstore be  $p_m^i$  and  $p_s^i$ , respectively. The price of any good a consumer buys from the superstore must satisfy  $p_s^i \leq p_m^i + \tau$ , and we assume this condition holds for all goods the superstore sells. (An equilibrium that satisfies

---

<sup>10</sup>This convexity is consistent with the findings of [Kosová and Lafontaine \(2008\)](#) for franchised chains. We abstract from the spatial dimension of the chain's expansion modeled by [Holmes \(2008\)](#) and [Jia \(forthcoming\)](#).

this condition is guaranteed below.) Given this condition, a consumer who makes a trip to the superstore will buy all  $n$  items offered by the superstore for the bundle price of

$$P_s \equiv \sum_{i=1}^n p_s^i. \quad (1)$$

A consumer chooses to shop at the superstore over purchasing all  $N$  lines at Mom and Pop stores if and only if

$$\sigma + \tau + P_s + \sum_{i=n+1}^N (p_m^i + \tau) \leq \sum_{i=1}^N (p_m^i + \tau).$$

This defines a threshold value for aversion:

$$\bar{\sigma} \equiv (n-1)\tau + \sum_{i=1}^n p_m^i - P_s. \quad (2)$$

Consumers with aversion  $\sigma \leq \bar{\sigma}$  shop at the superstore, whereas consumers with  $\sigma > \bar{\sigma}$  “boycott” it. The superstore has some shoppers as long as  $\bar{\sigma} > 0$ , that is, as long as

$$P_s \leq (n-1)\tau + \sum_{i=1}^n p_m^i \quad (3)$$

and the measure of consumers who shop at the superstore is given by

$$F_\sigma(\bar{\sigma}) = 1 - \exp(-\bar{\sigma}).$$

This measure depends, through  $\bar{\sigma}$ , on the price and size of the superstore’s bundle, prices at the Mom and Pop stores, and (for  $n > 1$ ) on consumers’ transportation cost per store.

### 3.3 Price Game

In the second stage, all  $N$  Mom and Pop stores in each location and the superstore simultaneously set prices, taking as given the superstore’s first-stage choices of  $n$  and  $k$ , and

anticipating consumers' reaction to their choices.

As noted earlier, we start by assuming that  $p_s^i \leq p_m^i + \tau$  for all lines and that  $\bar{\sigma} > 0$ , which guarantees positive demand at both Mom and Pop stores and the superstore. Later, we verify that these conditions are satisfied. Demand for each of the superstore's products is

$$x_s^i = F_\sigma(\bar{\sigma})$$

while each Mom and Pop store has demand

$$x_m^i = 1 - F_\sigma(\bar{\sigma})$$

in locations with superstore presence.<sup>11</sup>

The  $i$ th Mom and Pop store solves

$$\max_{p_m^i} \pi_m^i = (p_m^i - c_m^i) \cdot x_m^i(p_m^i, p_m^{-i}; P_s, n).$$

Assuming an interior solution, the first-order condition for the  $i$ th Mom and Pop store is

$$(1 - F_\sigma(\bar{\sigma})) - (p_m^i - c_m^i) \cdot f_\sigma(\bar{\sigma}) \frac{\partial \bar{\sigma}}{\partial p_m^i} = 0$$

which solves to the Mom and Pop's best-response function

$$p_m^{i*}(P_s) = c_m^i + 1. \tag{4}$$

In other words, in markets where a superstore competes and sets a price not too high (setting a price too high effectively prices the superstore out of the market), Mom and Pop stores'

---

<sup>11</sup>In locations—lines in which the superstore does not compete, the Mom and Pop stores are local monopolists. Their profit-maximizing prices are independent of the other stores' choices and equal to consumers' reservation price less their transportation cost. We implicitly assume that consumers' reservation price does not impose a binding constraint on strategically price-setting stores.

dominant strategy is to set their prices equal to  $c_m^i + 1$ . The Mom and Pop stores' fixed markup of 1 over marginal cost is a feature of the exponential distribution of big-box aversion (1 is the value of the exponential distribution function's rate parameter) and guarantees the chain's second-stage profit is uniquely identified for each first-stage choice of  $n$  and  $k$ . In a more general case the Mom and Pop stores' best-response function would depend not only on the superstore's prices but also on the other Mom and Pop stores' prices, in which case multiple equilibria are possible.

At the same time, the superstore chooses its vector of prices to maximize profit, taking all the Mom and Pop stores' prices affecting demand (and its own prior choices of scale and scope) as given. To save notation, variables without subscripts refer to the superstore. The superstore maximizes profit at a representative store,

$$\max_{p^1, \dots, p^n} \pi = \sum_{i=1}^n \left( x^i(p^1, \dots, p^n; p_m^1, \dots, p_m^n, n, k) \cdot \left( p^i - \frac{w(kx^i(\cdot))}{kx^i(\cdot)} \right) \right).$$

From the consumer's problem we know that as long as the superstore's individual prices satisfy conditions for an interior solution, only the price of the bundle,  $P$ , affects demand, so we can simplify the superstore's optimization problem to:

$$\max_P \pi = F_\sigma(\bar{\sigma}(P, p_m^1, \dots, p_m^n)) \cdot P - \frac{n \cdot w(k \cdot F_\sigma(\bar{\sigma}(\cdot)))}{k}.$$

Equivalently, we can work with the inverse demand function  $P = P(x)$  and write the superstore's optimization problem as a quantity choice:

$$x^* = \arg \max_x \pi = x \cdot P(x) - \frac{nw(kx)}{k} \tag{5}$$

where  $x = x^i = x^j$  for all  $i, j$ . The optimal bundle price can be found through the inverse demand function:

$$P^* = P(x^*) = \ln(1 - x^*) + (n - 1)\tau + \sum_{i=1}^n p_m^i. \tag{6}$$

The first-order condition for problem (5) is

$$\begin{aligned} \pi_x(x^*; p_m^{1*}, \dots, p_m^{n*}) &\equiv \frac{\partial \pi}{\partial x} \\ &= \ln(1 - x^*) + (n - 1)\tau + \sum_{i=1}^n p_m^i - \frac{ax^*}{1 - x^*} - nw'(kx^*) = 0. \end{aligned} \quad (7)$$

Equation (7) defines the superstore's best response bundle price,  $P(x^*)$  by way of quantity  $x^* = x(p_m^1, \dots, p_m^n, \tau, n, k)$ .

**Lemma 1** (Existence of Nash Equilibrium). *An interior Nash Equilibrium of the second-stage price game exists and is generically unique.*

All proofs are in Appendix A.

Lemma 1 establishes that there is a bundle price  $P^* = P(x^*)$  which maximizes the superstore's profit and satisfies  $\bar{\sigma} > 0$  and  $p_s^i \leq p_m^i + \tau$  for all  $i \leq n$ . The Nash equilibrium of this game uniquely identifies the superstore's second-stage profit,  $\pi^*$ , for given first stage choices  $n$  and  $k$ . Generically, the bundle price is unique but there is a continuum of solutions for the full vector of the superstore's prices: any vector that sums to  $P^*$  whose elements satisfy the condition for positive demand is a Nash equilibrium of this game. A focal price vector is one in which all markups are equal, as in Result 2 of Bliss (1988), but that is not necessary.

The properties of the Nash equilibrium are summarized in Figure 4 depicting the two best-response functions. The Mom and Pop's price,  $p_m^i$ , is on the  $x$ -axis, and the quantity sold by the superstore is on the  $y$ -axis. An exogenous increase in the number of stores operated by the chain,  $k$ , increases its volume of sales and thereby reduces marginal cost for each item sold by the superstore, drawing more shoppers to it. An exogenous increase in the number of products sold,  $n$ , also draws more shoppers because of the one-stop shopping effect. A higher transport cost,  $\tau$ , or higher cost for the Mom and Pop,  $c_m^i$ , likewise increase the benefit from to one-stop shopping and brings more shoppers to the superstore.

The existence proof guarantees that, in the second stage (i.e., net of sunk costs), each

superstore is profitable at any value of  $n$  and  $k$ . The chain sets  $n$  and  $k$  in the first stage, balancing the marginal benefit of each — the higher operating profits — against the marginal cost associated with the increased complexity of its logistics, distribution, and management problem.

### 3.4 Superstore's Scale and Scope

In the first stage the superstore chooses the number of product lines it carries,  $n$ , and the number of stores it operates,  $k$ . To conserve on notation, from here on we assume that costs for all  $N$  goods are symmetric. The first-stage profit function is

$$\Pi(n, k) = k \cdot \pi(x^*(n, k; c_m, \tau), n, k; c_m, \tau) - \frac{d(n, k)}{\delta}$$

where  $x^*(n, k; c_m, \tau)$  is the second-stage Nash equilibrium quantity of a line sold by the superstore. Our main result, stated formally below, is that when transportation cost is sufficiently high, the superstore's scale and scope are complements. That is, an increase in  $k$  ( $n$ ), resulting from some exogenous shock, will lead to a reinforcing increase in  $n$  ( $k$ ).

To state more precisely what we mean by sufficiently high transportation cost, define

$$\bar{\tau} \equiv \frac{w(x^*(1, 1; \cdot)) + \frac{1}{\delta}R}{x^*(1, 1; \cdot)} - (c_m + 1), \quad (8)$$

where  $R$  is the upper bound on  $\frac{\partial^2 d(n, k)}{\partial n \partial k}$ . When  $\tau > \bar{\tau}(R, c_m, \delta)$  the cost of offering a unit of an additional product is lower than the cost, inclusive of transportation cost, of acquiring a unit of the same product at a Mom and Pop store. In other words, when transportation cost is high enough, there is a positive surplus that the superstore can potentially capture by adding another line.

Our main result is stated below:

**Result 1** (Supermodularity).  $\Pi(n, k; c_m, \tau, \delta)$  is supermodular in  $(n, k; c_m, \tau, \delta)$  on the domain  $\tau > \bar{\tau}(R, c_m, \delta)$ .

Result 1 implies that for sufficiently high transportation cost,  $\tau$ , the superstore's optimal choices of number of lines,  $n$ , and number of stores,  $k$ , are complementary. Moreover, both move together (increasing or decreasing) with changes in the economic environment.<sup>12</sup>

The domain restriction captures the importance of the one-stop shopping effect to the complementarity of  $n$  and  $k$ . Benefits from one-stop shopping derive from the consumer having to make fewer trips to purchase a given bundle of goods. Thus, one-stop shopping benefits increase with transportation cost,  $\tau$ . If  $\tau$  is sufficiently high, the surplus the superstore can extract when it increases the number of lines it carries is sufficient to offset its increased coordination cost.

The superstore's scale economies also contribute to the complementarity of lines and stores. When the chain adds stores, it increases the number of locations in which it earns a profit. The degree to which profit increases with the number of stores in the chain depends on the number of lines sold through the one-stop shopping effect, as described above. The increase in stores also increases the chain's total volume of sales for each line which lowers the marginal cost of purchasing for *all* lines. Thus the marginal gain of increasing the number of lines sold also increases with the number of stores in the chain.

The broader implication of this result, which we can take to the data, is that  $n$  and  $k$  move together in response to changes in the economic environment. A change such as an increase in the cost of Mom and Pop stores or an improvement in technology affects both the chain's optimal scale and its optimal scope in the same way, and these increases in scale and scope reinforce and amplify each other.

To see the intuition for this, consider an increase in the Mom and Pop stores' cost parameter  $c_m$ . Such an increase prompts Mom and Pop stores to increase their prices,

---

<sup>12</sup>Available technology itself may be endogenous; see [Basker and Van \(2008\)](#). The proof can be easily modified to accommodate this change.

shifting out demand at the superstore’s existing locations. (The effect is the same even if only one, or a subset, of Mom and Pop stores are affected by the cost increase.) The higher profit at each store in the chain induces the chain to expand, adding more stores, which lowers marginal cost (through economies of scale) and increases profit further. In addition, higher sales of each product raise the profitability of each product and triggers additions to the chain’s selection of products.

A similar argument shows that an increase in  $\tau$ , consumers’ transportation cost per store, increase both the chains’ optimal scale and its optimal scope by shifting out demand at each store in the chain and thereby increasing the profitability of each sale. On the technology side, bar codes and other technological innovations which increase the optimal size of a store (Holmes, 2001) likewise increase the chain’s incentive to grow, leading not only to larger superstores but also to larger *chains* of superstores.

## 4 Data

To test the model, we use micro data from the Census of Retail Trade (CRT) for the years 1977, 1982, 1987, 1992, 1997, and 2002.<sup>13</sup> Detailed Economic Census forms, from which our data are compiled, are mailed to every chain store, as well as to a sample of single-unit retailers with more than 10 employees.<sup>14</sup> Data on stores that do not receive (or do not return) forms comes from administrative records.

The unit of observation in the CRT is the establishment, or store. We are able to aggregate the store-level information to the firm level by using firm identifiers for each store. These firm identifiers also allow us to track ownership over time. We limit our analysis to firms with establishments in general merchandising, SIC 53 (or NAICS 452). We further

---

<sup>13</sup>See Foster, Haltiwanger, and Krizan (2006) for more details on the CRT.

<sup>14</sup>The Census sends forms to all firms operating at least two establishments even if the firm operates only a single retail store (other establishments operated by the firm may include manufacturing plants or wholesale establishments).

limit our analysis to stores that received Census forms numbered 5301 or 5302 (in 1982–1997) or 45201 or 45202 (in 2002).<sup>15</sup> These forms are designated for general-merchandise stores, although a small fraction of general-merchandise stores receive other forms (for example, forms intended for apparel stores or supermarkets).

The other major piece of information we obtain from the CRT is the number of distinct product lines sold by each store. Each form includes a list of possible product lines, such as “toys, hobby goods, and games” and “major household appliances.” There are two types of product lines: “broad” lines and “detailed” lines, which provide a detailed breakdown of the broad lines. Table B-1 lists the broad categories of goods, along with the years they are included on the forms and the number of detailed lines associated with each. The broad line “groceries,” for example, includes up to nine detailed lines: meat, fish, and poultry; fresh and prepackaged produce; frozen foods; dairy products; bakery products; deli items; soft drinks; candy; and all other foods. (This level of detail is available only for 2002, however, and only for stores receiving form 45202; all other forms include the broad line “groceries” without the detailed breakdown.)

A product line encompasses many substitutes, with minimal substitution *across* product lines. For example, there are many possible substitute outfits within the broad line women’s apparel, but little substitution between women’s apparel and men’s apparel. Substitution across detailed lines within a broad line, such as women’s suits and women’s slacks, is more likely, but even here the definitions are, in general, broad enough so that a product’s most common substitutes will lie within the same detailed product line.

We take several steps to adjust for increases over time in the number of lines listed on the Census forms. Some of the increase in reported product lines are due to the increased detail on Census forms and some to the growing array of products carried by general-merchandise stores. To adjust for changes in the composition and definition of product lines over time,

---

<sup>15</sup>We do not restrict observations by form in 1977 because form information is not available for that year.

we created a longitudinally-consistent concordance of product lines. (The full concordance is available upon request.) For example, paint was included as a detailed line under the broad category of building materials in 1977 and 1982 but was listed as a separate broad line beginning with the 1987 forms; we recast it as a detailed line in the later years. As another example, until 1987 the women’s apparel broad line included a detailed line for girls’ and infants’ apparel, and the broad line of men’s apparel included boys’ apparel as a detailed line. Starting in 1992, Census forms included children’s apparel as a separate broad line. We reclassified all boys’ and girls’ apparel from the earlier years, listed as part of the adult apparel broad lines, as children’s apparel. This was not always possible, however. Pet and pet supplies, for example, were not listed on any general-merchandise form until 1987 except under the broad line of miscellaneous merchandise. Rather than throw away this information, we allowed a new line to start in 1987. We control for  $\text{year} \times \text{fraction of establishments}$  in each firm receiving each form in our regression analysis to capture this and a number of other changes over time.

In most cases, firms report product-line sales for all their stores. However, we do control for a firm’s **coverage**: the share of its stores reporting lines data. Average coverage is very close to 1 (we omit stores that do not report any lines data from this calculation). An implicit assumption is that the reporting stores are representative of the chain as a whole. We impute lines in only a few special circumstances. First, if a store reports sales under a broad product category but does not indicate which detailed line(s) it sells, we add one to its count of detailed lines. Similarly, if a store reports selling a detailed product line we add the relevant broad line to its count of lines. Next, if a store reports selling a product line that does not appear on the form it received (a “write in”) we add both a detailed line and a broad line to its count of lines. Finally, we calculate the average number of lines carried by stores within a chain; we do this calculation separately for broad lines and detailed lines. This measure is numerically identical to a weighted average of lines in which each line is weighted by the fraction of stores in the chain that carry that line.

Our handling of both store counts and line counts is conservative in the sense that we avoid making assumptions that would create a spurious relationship between the number of stores in a chain and the number of lines it carries. Consider two chains, one with 10 stores and one with 15 stores. The first chain carries lines A, B, and C (and only those lines) in each of its stores, so the number of products this chain carries is unambiguous: three. The second chain carries lines A, B, and C, in ten of its stores and lines A, B, and D (but not C) in the other five. Thus, although the number of lines carried by each stores in the second chain is 3, four unique products are represented in the chain’s inventory. By using the average number of products carried by each store — three — rather than the maximal number, we avoid counting more lines for larger chains simply because a single store might carry a different item. A third chain with 16 stores, of which 15 report line sales, carries lines A, B, and C in each of these 15 stores and line D in five stores is assigned a line count of three and one third, and coverage of 94%. Each time a single store carries one new line, the total number of lines the chain carries increases by  $\frac{1}{k}$ , where  $k$  is the number of stores in the chain. We have also tried alternative measures, including counting a line if and only if it is carried by at least a fixed fraction of stores in the chain. Results were qualitatively similar.

Table 3 provides summary statistics on the data used in the regressions. The data set is not a balanced panel, but represents a snapshot of the general-merchandise subsector in each of the six census years. A total of 17,431 observations represent 12,827 retailers; of these, roughly 14,000 are single-store retailer-years, and about 2,800 retailer-years are small chains (averaging 8.8 stores per chain). Only about 200 retailer-year observations have at least 100 stores, averaging 550 stores per chain. The total number of observations is smaller than the one reported in Table 1 for two reasons. First, as noted above, we omit firms that did not get the proper forms from our data set. Second, if none of the stores operated by a given retailer report any product line sales, we do not use that retailer in the data. This is much more likely to happen for single-store retailers, of which we keep only 18%. In contrast,

virtually all the large-chain observations from the CRT are included in our data.

The last column in Table 3 provides the share of firms (by size class) that are **dynamic**, that is, for which we observe some change in the number of lines sold and/or the number of stores operated. Dynamic firms exclude nearly 10,000 firms, the vast majority of them single-store retailers, that appear only in a single census. Because we include firm fixed effects in all of the regressions, observations of non-dynamic firms do not contribute to the estimation of the relationship between firm scale and scope. There are 2,951 dynamic retailers in the data, representing 7,555 observations. Almost all large chains (93%) are dynamic, as are 73% of small chains and 34% of single-store retailers.<sup>16</sup> The average dynamic firm in the sample has 19 stores and carries 15 broad lines and 22 detailed lines: it has almost twice as many stores as the average firm overall, and carries 20% more product lines.

## 5 Empirical Findings

We use a difference-in-difference specification to control for unobserved differences across firms and changes over time in Census forms and product line definitions and to isolate the within-firm relationship between number of products carried and number of stores operated. In our most general specification, we estimate

$$\ln(n_{rt}) = \sum_r \alpha_r + \sum_{ft} (\theta_t \cdot \phi_{rft}) + \beta \cdot \ln(k_{rt}) + \gamma \cdot \text{coverage}_{rt} + \varepsilon_{rt} \quad (9)$$

where  $n_{rt}$  is the number of (broad or detailed) lines of merchandise carried by retailer  $r$  in year  $t$ ,  $\alpha_r$  is a retailer fixed effect for each of the 12,827 retailers in the sample,  $\theta_t$  is a year fixed effect (for 1977, 1982, 1987, 1992, 1997, and 2002),  $\phi_{rft}$  the fraction of stores belonging to retailer  $r$  that received form  $f$  in year  $t$  (for forms 5301, 5302, 45201, and 45202), and  $k_{rt}$

---

<sup>16</sup>The set of dynamic firms is nearly invariant to whether we use detailed lines or broad lines; almost all firms with two or more observations are dynamic firms.

is the number of stores operated by retailer  $r$  in year  $t$ . The retailer fixed effects ensure that we identify the relationship between  $k$  and  $n$  using only within-firm variation over time; only dynamic firms contribute to the parameter identification. We include year  $\times$  form controls because, as noted earlier, different forms contain different counts of detailed and broad lines, and these change over time. The control variable **coverage** is the fraction of establishments within a firm for which valid line counts are obtained.

Our theoretical model predicts that  $\beta$ , the elasticity of the number product lines a retailer carries with respect to the number of stores in the chain, is positive. But we do not interpret estimates of  $\beta$  causally; in our model, both  $n$  and  $k$  are choice variables for each firm and their optimal values move with available technology ( $\delta$ ), consumers' transportation cost ( $\tau$ ), and rival stores' wholesale prices ( $c_m$ ), among other parameters. The fact that a retailer's choice of stores and lines is affected similarly (in the same direction) by changes in costs and technology parameters allows us to estimate the reduced-form relationship between them without taking a stand on the exact exogenous factors underlying the growth in both scale and scope over the past several decades.

Table 4 reports our estimates of  $\beta$ . The first three columns use the (log of) the count of broad lines on the LHS, and the last three use the count of detailed lines. All specifications control for year and retailer fixed effects; in each set of three, the second controls also for coverage, and the third controls also for differential form effects by year.

Estimates of  $\beta$  for broad lines fall in a tight range around 6.5% and are significant at least at the 5% level across specifications. One way to interpret these figures is to calculate how many broad lines a firm adds when it opens one additional store. Consider the average dynamic firm in the data, with approximately 19 stores and 15 broad lines. A change in economic conditions that leads a chain with this scale and scope to add one store will also increase its total number of lines by about 0.05. Since increasing the number of lines by 1 would mean adding a line to each of the stores in the chain, adding 0.05 lines is equivalent, approximately, to adding a line to one existing store.

We perform similar calculations in the top panel of Table 5 for each of the three firm sizes: single-store retailers, small chains, and large chains. We report the number of stores that are expected to add one line in conjunction with the opening of a new store; the direct competitive effect, measured as the total number of new markets in which the retailer is expected to compete; and the indirect competitive effect, measured as the total number of markets in which the retailer was already operating and where competitors face an inward shift of the demand function due to the one-stop shopping effect.

We use the coefficient estimates from the third column in Table 4 for this calculation, since that specification includes all relevant control variables. Consistently across firm sizes, one new store is correlated with one new line in an existing store. The lowest estimate is that 0.88 stores adopt a new broad line when a new store opens (for single-store retailers) and the highest estimate is that 1.39 stores adopt a new broad line when a new store opens (for a large chain). The range of the direct and indirect effects is slightly higher, with magnitudes increasing with the retailer's size. For example, the number of markets directly affected when a single-store retailer adds a store is about 13: twelve markets through the addition of a store, and approximately one more because the existing store adds one line. For a large chain the number of new markets is about 22: twenty one markets are directly affected by the new store, and one additional market is affected by the superstore that adds a product line. The indirect competitive effect associated with the addition of a store likewise increases with the retailer's size because it measures the number of markets in which the retailer already competes and which are subject to the one-stop shopping effect due to the addition of product lines by existing stores. For a single-store retailer who carries 12 broad lines, adding a broad lines in 0.88 stores indirectly affects  $0.88 \cdot 12$ , or 10.5, markets. Similarly, for a large chain carrying 21 product lines, adding another broad line in 1.39 stores indirectly affects  $1.39 \cdot 21$ , or 29, markets.

Estimates of  $\beta$  for detailed lines, in the last three columns of Table 4, range from 6.8–7.4%. The next set of numbers in Table 5 presents the number of stores that are expected

to add a detailed line in conjunction with adding one store based on the elasticity estimates from the last column in Table 4. A single-store retailer adds about 1.3 detailed lines in an existing store when it opens an additional store; a large chain adds a detailed line in about 2.6 existing stores with each new store opening. The direct and indirect competitive effects are similar in magnitude to the effects reported for broad lines.

An alternative way to illustrate the complementarities between scale and scope is to calculate how many stores a retailer of each size adds when it increases the number of lines it carries by one. The bottom two panels of Table 5 quantify the competitive effects of a retailer adding a broad and a detailed line, respectively. These effects increase dramatically with the chain's size. To see why, suppose a large chain carrying 20.7 lines in 555.4 stores responds to changes in its economic environment by adding one new broad line to each of its stores. The complementarity between scale and scope implies it also opens 383 new stores. The combination of new stores and new products adds many new markets in which the retailer now competes directly: in addition to the 555.4 new markets due to the product expansion, each of the new stores competes in 20.7 product markets, for a total of 8,483 new markets.

Put differently, adding a broad line throughout the chain is associated with increasing the number of markets in which a large chain competes by about 55% from a base of 11,496. Adding a product line is a much bigger undertaking in a large chain than in a small one, because it entails adding a line in each store in the chain. In addition, the larger product selection in the chain's existing stores increases the chains' competitive pressures in each of the markets in which it initially participated. A similar calculation shows that adding a detailed line increases the number of markets in which a large chain competes, on average, by 35%. For small chains and single-store general merchandisers, the competitive effects associated with adding a product line are smaller in absolute value but larger in relative terms. Adding a broad line to a stand-alone retailer, for example, is accompanied by doubling the number of stores (turning it into a small chain), which also doubles the number of markets

in which it competes. Likewise for a small chain: when it adds a broad line it also adds 8.2 stores, just shy of the number (8.9) it already operates.

We perform several robustness checks (not shown) of this analysis. First, to address the possibility that a single retailer may operate multiple chains, we estimated the elasticity of lines with respect to stores within a **sub-chain**, defined as the set of all stores within a chain that share the same industry.<sup>17</sup> The estimated elasticities are not sensitive to this change. Second, we check that the qualitative results are similar when we use levels rather than logs for both stores and lines. The levels results yield positive and statistically significant estimates for  $\beta$  across all specifications.

Overall, we conclude that the relationship between chain size and the scope of products carried is statistically strong and economically important. When chains expand, they typically grow in both scale and scope, adding a broad product line to one existing store with each new store they open. These results are conservative in the sense that they are identified using only *contemporaneous* changes in scale and scope (within five-year periods). Adding a product line throughout a large chain can take many years if it requires reconfiguring layouts and perhaps building new stores (for example, in the case of groceries when refrigerated facilities are needed); if a chain adds a line and subsequently adjusts its product offerings, or vice versa, we will not capture this effect.

## 6 Concluding Remarks

In this paper, we document and offer an explanation for the simultaneous rise in chain size and product offerings by general merchandisers over the last several decades. Analyzing

---

<sup>17</sup>Unlike tracking chains across census years, following these sub-chains over time is complicated by the SIC-to-NAICS switch in the 1997 CRT. (There are three 4-digit SIC codes (5311, 5331, 5399) and two 4-digit NAICS codes (4521, 4529) under general merchandise, and the correspondence between them is imperfect. For example, NAICS 4521 was assigned to 85% of establishments in SIC 5311; the other 15% were assigned NAICS code 4529. (See <http://www.census.gov/epcd/ec97brdg/> for details.) In the absence of detailed information on both SIC and NAICS at the establishment level, we map SIC 5311 to NAICS 4521.

firm-level data spanning twenty five years and capturing both the scale and scope of general-merchandise firms, we show that the chains that have grown the most over this period have also expanded into the most new product markets.

Our model explains this relationship with an interaction of economies of scale and demand-driven gains from scope due to consumer preference for one-stop shopping. A retailer's optimal store size may increase if consumers' preference for one-stop shopping shifts due to changes in the composition of the labor-force or other changes in the value of time or the cost of gasoline. Technological innovations, such as the introduction of Radio Frequency Identification (RFID) technology in the distribution process, or cost reductions due to trade liberalization, increase a chain's optimal scale. The interaction between the two amplifies the effect of any one of these forces on both scale and scope.

As large general-merchandise chains add more lines, and more stores, they compete more vigorously with an increasing number of small businesses. Our results show that the addition of one product line to the offerings of the average large chain affects 19,980 markets: about half due to the opening of new stores and the rest due to increased one-stop shopping benefits at existing superstores. A consumer who now goes to the superstore to buy groceries also buys shoes and sporting goods, reducing demand at the Mom and Pop shoe and sporting-goods stores in town.

Empirical studies of the competitive effects of large general-merchandise chains such as Wal-Mart and Kmart have focused on one easy-to-measure outcome: store closings in the immediate vicinity of the new superstores. [Basker \(2005\)](#) estimates that each new Wal-Mart store accounts for, on average, the failure of 4.7 small and medium stores (with fewer than 100 employees). Over the course of the quarter century from 1977 to 2002 a total of 2,632 new Wal-Mart stores opened ([Basker and Noel, 2007](#)), which, by this estimate, have caused the closure of over 12,000 competing stores. On a similar scale, [Jia \(forthcoming\)](#) also finds that over half of the decline in single-store general merchandisers in the U.S. between 1987 and 1997 was due to the expansion of Wal-Mart and Kmart. But these numbers miss the

reduction in business at existing stores that do not fail and indirect effects on other markets due to the feedback from scale to scope and vice versa. All told, Census of Retail Trade figures show a net decrease of about 400,000, or 30%, in the number of retail establishments (excluding restaurants) with fewer than 100 employees between 1977 and 2002.

The competitive pressure exerted by these large chains is a testament to their greater efficiency. Other studies also find a robust correlation between a retailer's scale and scope and its efficiency. In the U.S., evidence from the Census of Retail Trade confirms that large (and expanding) chains throughout the retail sector are much more efficient than retailers operating a single store (Foster, Haltiwanger, and Krizan, 2006). In the U.K., Haskel and Sadun (2007, forthcoming) find that entry regulation that increased the cost of opening large stores reduced chain-level TFP growth, particularly in the "non-specialized" sector which includes both department stores and supermarkets. Lagakos (2008) argues that differences in retail productivity between the U.S. and less-developed countries are due almost entirely to differences in retail formats: in the U.S. most retail commerce occurs in "modern" establishments, whereas in most developing countries "traditional" Mom and Pop formats dominate.

As we argue in this paper, the higher efficiency of the bigger retailers reflects economies of scale in addition to the use of better technology. But a firm's level of investment in technology is itself a choice variable, and larger retailers consistently make bigger investments. Evidence from Beck, Grajek, and Wey (2007) and Doms, Jarmin, and Klimek (2004) supports the view that larger firms invest more heavily in technology, specifically in IT, both in the U.S. and abroad. Basker (2007) calculates that between 1982 and 2002 the productivity growth rate at Wal-Mart, a major investor in technology, was three times higher than the rest of the general-merchandise subsector's. American chains' ability to expand with few regulatory restrictions in both product space and across a large geographical area may well explain both their higher investment levels in technology and their higher efficiency levels compared with chains in other countries (Haskel, Jarmin, Motohashi, and Sadun, 2007).

While consumers value one-stop shopping and the lower prices that have come with increased efficiency, larger superstores have engendered increased opposition in many communities. One reason for this opposition is the sheer pervasiveness of the largest chains and the growing numbers of consumers exposed to them. The list of concerns varies from town to town and neighborhood to neighborhood, but it is superstores' competitive edge that makes them impossible to ignore. As they grow further in both scope and scale, they will bring about even more opposition alongside efficiency gains.

# A Proofs

*Proof of Lemma 1.* The superstore's best response function  $x^*(p_m^{1*}, \dots, p_m^{n*})$  is defined by the first order condition  $\pi_x(x^*; p_m^{1*}, \dots, p_m^{n*}) = 0$ .

First we show that  $x^*(p_m^{1*}, \dots, p_m^{n*})$  is defined for all  $n \geq 1, k \geq 1$ . By inspection of Equation (7), we see that for all  $n \geq 1, k \geq 1$ :

- $\pi_x(\cdot)$  is continuous, differentiable over  $x \in (0, 1)$ ;
- $\pi_x(0) = 0 + (n-1)\tau + \sum_{i=1}^n p_m^i - 0 - nw'(0)$ . We assume  $w'(0) < c_m^i$  for all  $i$ . Therefore  $\pi_x(0) > 0$  when  $p_m^i = c_m^i + 1$  and by continuity of  $\pi_x(\cdot)$ ,  $\pi_x(0) > 0$  for some values of  $p_m^i < c_m^i + 1$  as well;
- $\lim_{x \rightarrow 1} \pi_x(x^*) = -\infty$ .

By the Intermediate Value Theorem, there exists an  $x^* \in (0, 1)$  such that  $\pi_x(x^*) = 0$ . Furthermore,  $\pi_{xx}(x^*) < 0$  is the second-order condition for the superstore's optimization problem which is satisfied by the assumption that marginal cost is flatter than marginal revenue at  $x^*$ .

Although there may be multiple local optima satisfying both the first- and second-order conditions, generically one of them must dominate globally. This guarantees that the superstore's best response,  $x^*(p_m^{1*}, \dots, p_m^{n*})$ , is generically single-valued (although it need not be continuous). Therefore, the Nash equilibrium exists and is defined by  $p_m^{i*} = c_m^i + 1$  for all  $i$  and  $P(x^*(c_m^{1*} + 1, \dots, c_m^{n*} + 1))$ .

The fact that  $x^* > 0$  guarantees that  $\bar{\sigma} > 0$ , which was one of our starting assumptions. It also guarantees that  $P^* < \sum_{i=1}^n p_m^i + (n-1)\tau$ , which means that there exists a continuum of price vectors  $\{p^1, \dots, p^n\}$  satisfying  $p_s^i \leq p_m^i + \tau \quad \forall i$ . Moreover, second-stage profit per store is increasing in the number of products priced below  $p_m^i + \tau$ :

$$\frac{\partial \pi}{\partial n} = \frac{1}{k} \cdot (kx(\tau + p_m^n) - w(kx)) > 0,$$

so the superstore cannot be optimizing if it prices some products above this threshold.

The superstore's best-response function,  $x^* = x(p_m^1, \dots, p_m^n, \tau, a, n, k)$ , has the following properties:

$$\begin{pmatrix} \frac{\partial x}{\partial p_m^i} \\ \frac{\partial x}{\partial \tau} \\ \frac{\partial x}{\partial n} \\ \frac{\partial x}{\partial k} \end{pmatrix} = -\frac{1}{\pi_{xx}} \begin{pmatrix} \pi_{xp_m^i} \\ \pi_{x\tau} \\ \pi_{xn} \\ \pi_{xk} \end{pmatrix} = -\frac{1}{\pi_{xx}} \begin{pmatrix} 1 > 0 \\ n-1 \geq 0 \\ p_m^n + \tau - w'(kx) > 0 \\ -x \cdot w''(kx) > 0 \end{pmatrix}.$$

□

*Proof of Result 1.* Supermodularity requires that  $\Pi$  has increasing differences in  $(n, k; c_m, \tau, \delta)$ , or equivalently, since  $\Pi$  is continuous and twice differentiable, that the cross-partial derivatives  $\frac{\partial^2 \Pi}{\partial k \partial n}$ ,  $\frac{\partial^2 \Pi}{\partial k \partial \delta}$ ,  $\frac{\partial^2 \Pi}{\partial k \partial \tau}$ ,  $\frac{\partial^2 \Pi}{\partial k \partial c_m}$ ,  $\frac{\partial^2 \Pi}{\partial n \partial \delta}$ ,  $\frac{\partial^2 \Pi}{\partial n \partial \tau}$ ,  $\frac{\partial^2 \Pi}{\partial n \partial c_m}$ , are all non-negative.

First, the cross-partial derivatives with respect to  $\delta$ :

$$\begin{aligned} \frac{\partial^2 \Pi}{\partial n \partial \delta} &= \frac{1}{\delta^2} \frac{\partial d}{\partial n} > 0 \\ \frac{\partial^2 \Pi}{\partial k \partial \delta} &= \frac{1}{\delta^2} \frac{\partial d}{\partial k} > 0. \end{aligned}$$

Using the inverse demand function for the chain, we calculate the cross partials for the first stage profit function with respect to the parameters  $c_m$  and  $\tau$ :<sup>18</sup>

$$\begin{aligned} \frac{\partial \Pi}{\partial c_m} &= nkx^* + k\pi_x \frac{\partial x^*}{\partial c_m} \\ \frac{\partial \Pi}{\partial \tau} &= (n-1)kx^* + k\pi_x \frac{\partial x^*}{\partial c_m}. \end{aligned}$$

Before we write out the expressions for the second cross-partial derivatives, it is useful

---

<sup>18</sup>We are implicitly assuming that all  $c_m^i$ 's are increasing together.

to note that the expressions are evaluated at the second stage Nash prices which means that  $\pi_x \equiv 0$  and  $\frac{d\pi_x}{dc_m} = 0$ . This allows us to simplify the cross partial derivatives as follows:

$$\begin{aligned}\frac{\partial^2 \Pi}{\partial n \partial c_m} &= kx^* + kn \frac{\partial x^*}{\partial n} > 0 \\ \frac{\partial^2 \Pi}{\partial k \partial c_m} &= nx^* + kn \frac{\partial x^*}{\partial k} > 0 \\ \frac{\partial^2 \Pi}{\partial n \partial \tau} &= kx^* + kn \frac{\partial x^*}{\partial n} > 0 \\ \frac{\partial^2 \Pi}{\partial k \partial \tau} &= (n-1) \left( x^* + k \frac{\partial x^*}{\partial k} \right) \geq 0.\end{aligned}$$

Finally, we show that  $\frac{\partial^2 \Pi}{\partial n \partial k} > 0$ . Applying the envelope theorem to the second-stage price game, the derivative of  $\Pi$  with respect to  $k$  is

$$\frac{\partial \Pi}{\partial k} = \pi_s^* - \sum_{i=1}^n kx_s^i \frac{\partial}{\partial k} \left( \frac{w(kx_s^i)}{kx_s^i} \right) - \frac{1}{\delta} \frac{\partial d(n, k)}{\partial k} \quad (10)$$

where

$$\begin{aligned}\frac{\partial}{\partial n} \left( \sum_{i=1}^n kx^i \frac{\partial}{\partial k} \left( \frac{w(kx^i)}{kx^i} \right) \right) &= \frac{\partial}{\partial n} \left( \sum_{i=1}^n x^i \cdot \left( w'(kx^i) - \frac{w(kx^i)}{kx^i} \right) \right) \\ &= x^n \cdot \left( w'(kx^n) - \frac{w(kx^n)}{kx^n} \right) + \sum_{i=1}^n kxw''(kx^i) \frac{\partial x^i}{\partial n}\end{aligned}$$

so (imposing symmetry),

$$\frac{\partial^2 \Pi}{\partial k \partial n} = \frac{\partial \pi_s^*}{\partial n} + kx^* \cdot \left( \frac{w(kx^*)}{kx^*} - w'(kx^*) \right) - kn \frac{\partial x^*}{\partial n} w''(kx^*) - \frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k}. \quad (11)$$

The second term is positive by the concavity of  $w$  (average cost strictly exceeds marginal cost for  $kx > 0$ ). The third term is negative since  $w'' < 0$  (again by concavity of  $w$ ) and  $\frac{\partial x^*}{\partial n} = p_m^n + \tau - w'(kx) > 0$ , as shown earlier. Therefore a sufficient condition for Equation (11) to be positive is that

$$\frac{\partial \pi_s^*}{\partial n} - \frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k} > 0. \quad (12)$$

Incorporating the Mom and Pop stores' Nash price, the first term is

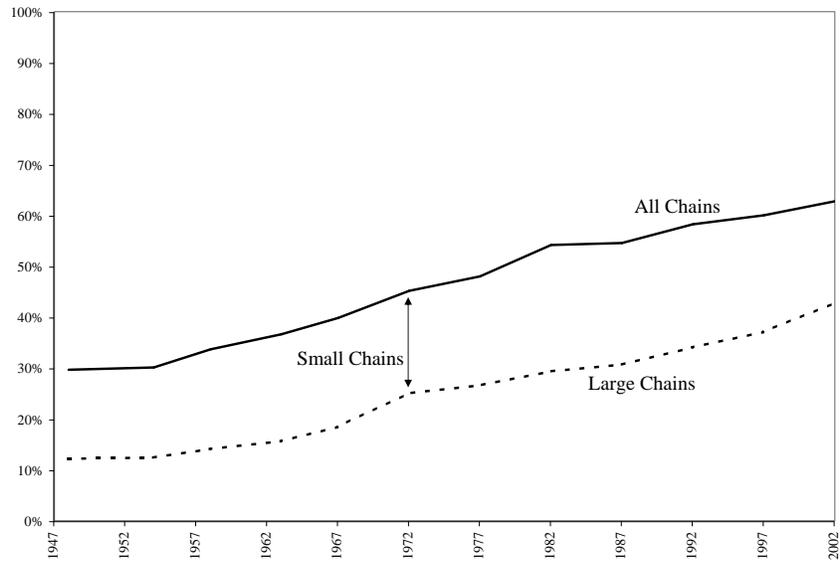
$$\frac{\partial \pi^*}{\partial n} = x^* \cdot \left( \tau + c_m + 1 - \frac{w(kx^*)}{kx^*} \right).$$

This term is everywhere positive, and is minimized when  $n = k = 1$  because  $x^*$  is increasing in  $n$ , and average cost,  $\frac{w(kx^*)}{kx^*}$ , is decreasing, so a sufficient condition is that Equation (12) holds at  $n = k = 1$ . This is guaranteed by the domain restriction  $\tau > \bar{\tau}(R, c_m, \delta)$ .  $\square$

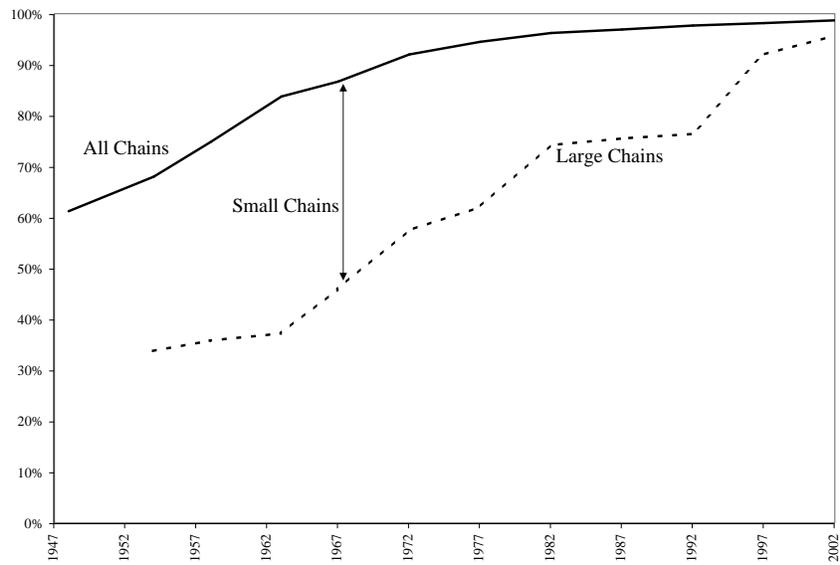
## References

- Bagwell, K., G. Ramey, and D. F. Spulber (1997) “Dynamic Retail Price and Investment Competition,” *RAND Journal of Economics*, 28(2), 207–227.
- Basker, E. (2005) “Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion,” *Review of Economics and Statistics*, 87(1), 174–183.
- (2007) “The Causes and Consequences of Wal-Mart’s Growth,” *Journal of Economic Perspectives*, 21(3), 177–198.
- Basker, E., and M. Noel (2007) “The Evolving Food Chain: Competitive Effects of Wal-Mart’s Expansion into the Supermarket Industry,” unpublished paper, University of Missouri.
- Basker, E., and P. H. Van (2008) “Wal-Mart as Catalyst to U.S.-China Trade,” unpublished paper, University of Missouri.
- Beck, J., M. Grajek, and C. Wey (2007) “Estimating Level Effects in Diffusion of a New Technology: Barcode Scanning at the Checkout Counter,” ESMT Working Paper 07-002.
- Betancourt, R., and D. Gautschi (1988) “The Economics of Retail Firms,” *Managerial and Decision Economics*, 9(2), 133–144.
- (1990) “Demand Complementarities, Household Production, and Retail Assortments,” *Marketing Science*, 9(2), 146–161.
- (1992) “The Demand for Retail Products and the Household Production Model,” *Journal of Economic Behavior and Organization*, 17(2), 257–275.
- Betancourt, R. R. (2004) *The Economics of Retailing and Distribution*. Edward Elgar, Cheltenham, UK, and Northampton, MA, USA.
- Bliss, C. (1988) “A Theory of Retail Pricing,” *Journal of Industrial Economics*, 26(4), 375–391.
- Bonné, J. (2004) “Big Store or Small? Why Consumers Care,” *MSNBC*, February 4, 2004, <http://www.msnbc.msn.com/id/4001039/4001039>.
- Doms, M. E., R. S. Jarmin, and S. D. Klimek (2004) “Information Technology Investment and Firm Performance in U.S. Retail Trade,” *Economics of Innovation and New Technology*, 13(7), 595–613.
- Foster, L., J. Haltiwanger, and C. J. Krizan (2006) “Market Selection, Reallocation and Restructuring in the U.S. Retail Trade Sector in the 1990s,” *Review of Economics and Statistics*, 88(4), 748–758.
- Gogoi, P. (2008) “Wal-Mart’s Crossover Strategy Still Works,” *Business Week*, March 7, 2008, [http://www.businessweek.com/bwdaily/dnflash/content/mar2008/db2008036\\_169954.htm](http://www.businessweek.com/bwdaily/dnflash/content/mar2008/db2008036_169954.htm).

- Haskel, J., R. Jarmin, K. Motohashi, and R. Sadun (2007) “Retail Market Structure and Dynamics: A Three Country Comparison of Japan, the UK and the US,” unpublished paper, LSE.
- Haskel, J., and R. Sadun (2007) “Entry Regulation and Productivity: Evidence from the UK Retail Sector,” unpublished paper, LSE.
- (forthcoming) “Entry, Exit and Labour Productivity in UK Retailing: Evidence from Micro Data,” in *Producer Dynamics: New Evidence from Micro Data*, ed. by J. B. Jensen, T. Dunne, and M. J. Roberts. University of Chicago Press.
- Holmes, T. (2001) “Bar Codes Lead to Frequent Deliveries and Superstores,” *RAND Journal of Economics*, 32(4), 708–725.
- (2008) “The Diffusion of Wal-Mart and Economies of Density,” National Bureau of Economic Research Working Paper 13783.
- Hotelling, H. (1929) “Stability in Competition,” *Economic Journal*, 39(153), 41–57.
- Jarmin, R. S., S. D. Klimek, and J. Miranda (forthcoming) “The Role of Retail Chains: National, Regional and Industry Results,” in *Producer Dynamics: New Evidence from Micro Data*, ed. by J. B. Jensen, T. Dunne, and M. J. Roberts. University of Chicago Press.
- Jia, P. (forthcoming) “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Industry,” *Econometrica*.
- Kosová, R., and F. Lafontaine (2008) “Survival and Growth in Retail and Service Industries: Evidence from Franchised Chains,” unpublished paper, University of Michigan.
- Lagakos, D. (2008) “Superstores or Mom and Pops? Market Size, Technology Adoption and TFP Differences,” unpublished paper, UCLA.
- Office of Management and Budget (1987) *Standard Industrial Classification Manual 1987*. Government Printing Office, United States of America.
- (1998) *North American Industry Classification System*. Bernan Press, United States of America.
- Pew Research Center for the People and the Press (2005) “Wal-Mart: A Good Place to Shop but Some Critics Too,” *Pew Research Center for the People and the Press*, December 15, 2005, <http://people-press.org/reports/display.php3?ReportID=265>.
- Ross, T. W. (1984) “Winners and Losers under the Robinson-Patman Act,” *Journal of Law and Economics*, 27(2), 243–271.
- Wal-Mart Stores, Inc. (1994) *Wal-Mart Annual Report*. Bentonville, AR.



(a) All Retail



(b) General Merchandise

Figure 1. All Chains' and Large Chains' Share of Dollar Sales  
 Source: Published data from Census of Business and Census of Retail Trade

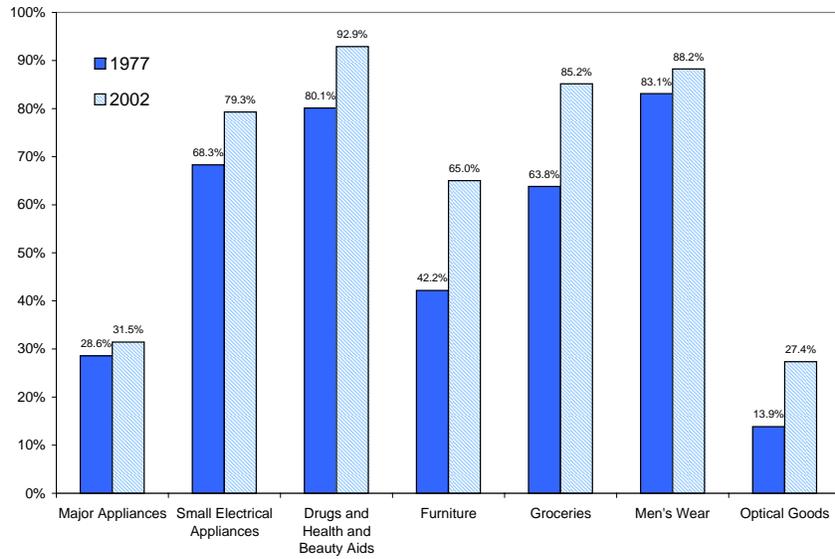


Figure 2. Share of General-Merchandise Establishments Carrying Selected Products  
Source: Census of Retail Trade

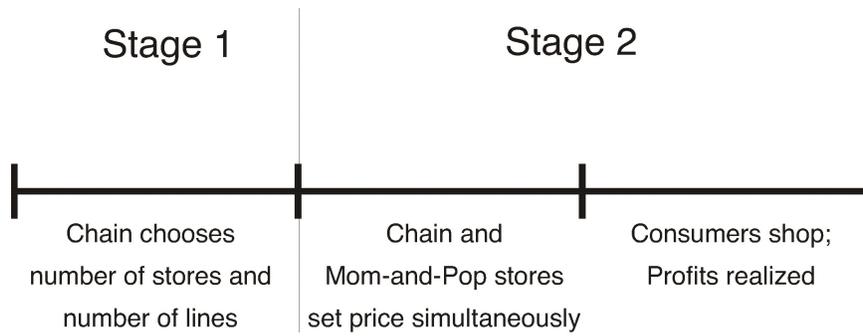


Figure 3. Timing of Model

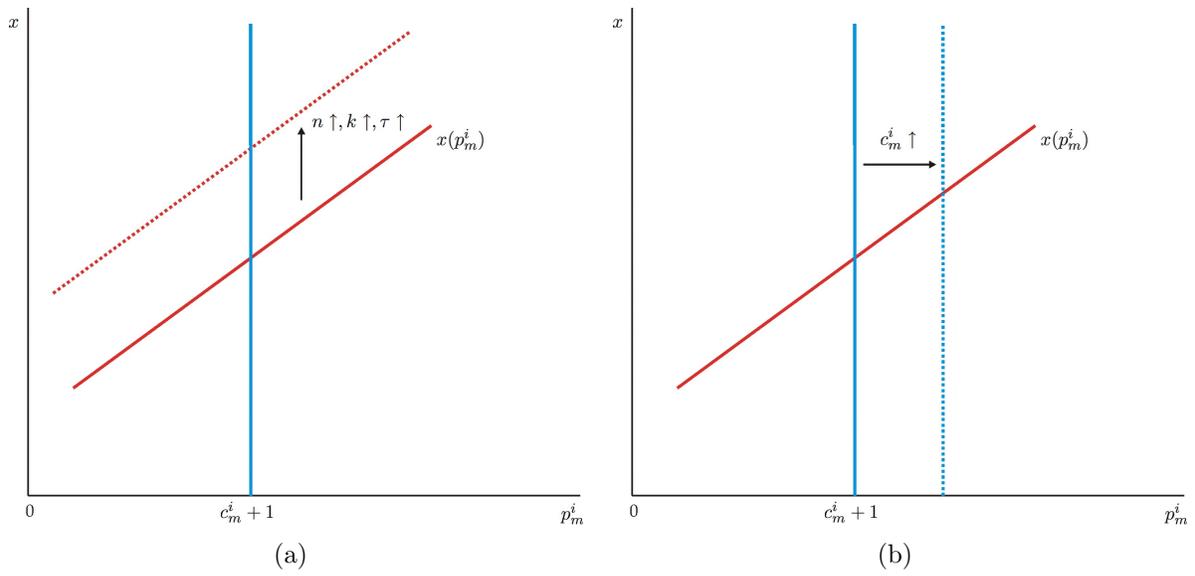


Figure 4. Comparative Statics of the Second-Stage Nash Equilibrium

Table 1. The General-Merchandise Subsector over Time

Year	All Firms	Single Store	Small Chain	Large Chain <sup>a</sup>	Total Sales <sup>b</sup>
1977	16,098	14,897	1,166	35	277,023,735
1982	13,344	12,386	918	40	222,424,639
1987	12,917	12,093	786	38	286,869,671
1992	10,264	9,660	565	39	305,466,876
1997	10,373	9,933	398	42	365,277,105
2002	9,467	9,150	285	32	433,068,999

Source: Census of Retail Trade, 1977–2002

<sup>a</sup> Large chains have 101 or more stores in 1977, 100 or more from 1982<sup>b</sup> Thousands of 2002 dollars

Table 2. Retail Sector vs. General-Merchandise Subsector, 2002

	Retail <sup>a</sup>	GM <sup>b</sup>	GM/Retail
Retail Firms (000)	727	9	0.013
Stores (000)	962	36	0.037
Sales (000,000,000\$)	2,975	433	0.146
Fraction of Stores in Chains	0.38	0.78	2.061
Fraction of Stores in Large Chains <sup>c</sup>	0.23	0.72	3.101

Source: Census of Retail Trade, 2002

<sup>a</sup> NAICS 44–45<sup>b</sup> NAICS 452<sup>c</sup> A large chain is a chain operating 100 or more stores

Table 3. Micro Data Summary Statistics: Variable Means

Chain Size	Observations	Stores per Firm	Broad Lines	Detailed Lines	Share Dynamic <sup>a</sup>
Single Store	14,437	1.0 (0.0)	12.0 (6.3)	16.6 (10.8)	0.34 (0.48)
Small Chain	2,784	8.9 (14.9)	14.8 (6.4)	23.0 (12.9)	0.73 (0.44)
Large Chain	210	555.4 (777.6)	20.7 (6.6)	35.0 (16.9)	0.93 (0.26)
All	17,431	8.9	12.5	17.9	0.41

Standard deviations in parentheses

<sup>a</sup> Dynamic firms are firms for which we observe changes in number of stores and/or number of broad lines. The share is extremely similar if we use the number of detailed lines in place of the number of broad lines or if we use the share of firms that appear at least twice.

Table 4. Regressions Results

LHS	ln(Broad Lines)			ln(Detailed Lines)		
ln(Stores)	0.0632*** (0.0204)	0.0674*** (0.0203)	0.0672*** (0.0203)	0.0681** (0.0272)	0.0737*** (0.0273)	0.0731*** (0.0273)
Chain FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Coverage		✓	✓		✓	✓
Form $\times$ Year			✓			✓

Each regression has 17,431 observations

Robust standard errors in parentheses (clustered by chain).

\*\* significant at 5%; \*\*\* significant at 1%.

Table 5. Economic Implications by Retailer Size

	Single Store	Small Chain	Large Chain
Effect associated with increasing number of stores by one:			
Number of stores that add a broad line	0.88	1.00	1.39
Direct competitive effect <sup>a</sup>	12.84	15.83	22.09
Indirect competitive effect <sup>b</sup>	10.54	14.80	28.79
Number of stores that add a detailed line			
Number of stores that add a detailed line	1.33	1.68	2.56
Direct competitive effect <sup>a</sup>	17.95	24.68	37.53
Indirect competitive effect <sup>b</sup>	22.17	38.70	89.40

Effect associated with increasing number of broad lines by one:

Number of new stores	1.11	8.20	383.0
Direct competitive effect <sup>a</sup>	14.22	130.5	8,483.2
Indirect competitive effect <sup>b</sup>	11.95	131.2	11,496.4

Effect associated with increasing number of detailed lines by one:

Number of new stores	0.78	5.11	214.5
Direct competitive effect <sup>a</sup>	13.96	126.5	8,058.3
Indirect competitive effect <sup>b</sup>	16.62	203.5	19,423.5

<sup>a</sup> Number of new markets in which retailer competes

<sup>b</sup> Number of existing markets in which competitors' residual demand shifts inwards

Table B-1. Product Lines

<b>Broad Line Description</b>	<b>Years</b>	<b>Details<sup>a</sup></b>
Groceries	All	9
Meals, Snacks, and Nonalcoholic Beverages for Immediate Consumption	All	1
Packaged Liquor, Wine, and Beer	All	3
Tobacco Products and Accessories	All	1
Drugs and Health and Beauty Aids	All	6
Soaps, Detergents, and Household Cleaners	1987–2002	1
Paper and Related Products	1987–2002	1
Men's Apparel	All	11
Women's Apparel	All	13
Children's Apparel	All	4
Footwear	All	5
Curtains, Draperies, and Domestic	All	2
Major Household Appliances	All	3
Small Electrical Appliances	All	1
Televisions, VCRs, and Videotapes	All	2
Audio Equipment and Music	All	3
Furniture	All	4
Floor Coverings	All	3
Computer Hardware and Software	All	2
Kitchenware and Home Furnishings	All	4
Jewelry	All	2
Optical Goods (Including Eyeglasses, and Telescopes)	All	1
Sporting Goods (Including Bicycles and Guns)	All	7
Hardware, Tools, Plumbing, and Electrical Equipment and Accessories	All	1
Lawn and Garden Equipment and Supplies	All	4
Building Materials, Paint, and Home Improvement Equipment and Supplies	All	1
Automotive Supplies	All	3
Automotive Fuels	All	1
Household Fuels	All	1
Pets, Pet Foods, and Pet Supplies	1987–2002	1
Photographic Equipment and Supplies	All	1
Toys (Including Games and Crafts)	All	2
Sewing, Knitting, and Needlework Goods	All	1
Stationary, School, and Office Supplies	All	2
Luggage and Leather Goods	All	1
Office Equipment	1987–2002	2
Souvenirs and Novelty Items, Including Seasonal Decorations	1997–2002	2
Books, Magazines, and Newspapers	All	2
Miscellaneous Merchandise, Not Elsewhere Classified	All	3
Non-merchandise Receipts	All	8

<sup>a</sup> Maximum number of detailed lines associated with each broad line.