

Supersize It: The Growth of Retail Chains and the Rise of the “Big-Box” Retail Format*

Emek Basker
University of Missouri

Shawn Klimek
US Census Bureau

Pham Hoang Van
Baylor University

September 2011

Abstract

This paper documents and explains the recent rise of “big-box” general merchandisers. Data from the Census of Retail Trade for 1977–2007 show that general-merchandise chains grew much faster than specialist retail chains, and that general merchandisers that added the most stores also made the biggest increases to their product offerings. We explain these facts with a stylized model in which a retailer’s scale economies interact with consumer gains from one-stop shopping to generate a complementarity between a retailer’s scale and scope.

JEL Codes: L11, L25, L81

Keywords: Retail, chain, big box, superstore, economies of scale, general merchandise, one-stop shopping

*Author contact: emek@missouri.edu, shawn.d.klimek@census.gov, or van_pham@baylor.edu. Any 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. We thank the coeditor, an anonymous referee, Saku Aura, Jonathan Beck, Roger Betancourt, Tom Davidoff, Lucia Foster, David Gautschi, John Haltiwanger, Tom Holmes, Tom Hubbard, Bob Hunt, Ron Jarmin, Renáta Kosová, Julia Lane, Michael Lynch, David Mandy, Guy Michaels, Tom Mroz, Leonard Nakamura, Kevin Tsui, Henry Wan, and seminar participants at Berkeley (Haas), Census Bureau, Clemson, Delaware, Department of Justice, Federal Reserve Bank of Philadelphia, Federal Trade Commission, George Washington University, IUPUI, London Business School, Missouri, Maryland, Washington University in St. Louis, the 2006 “Advances in the Empirical Analysis of Retailing” workshop at WZ-Berlin, the 2007 AEA (Chicago), the 2007 IIOC (Savannah), the 2008 RPI Mini-Conference on Marketing and Innovation and the 2010 Comparative Analysis of Enterprise Data (London) 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

This paper is motivated by two observations: the growth of retail chains, and the increased diversity of products sold in many chains. Examples of the latter include groceries sold at gas stations and drugstores, coffee shops incorporated into many bookstores, combination fitness centers/spas, and the addition of banks and pharmacies in many supermarkets. Another example is the Canadian bookstore chain Indio, which sells flowers and wine, among other offerings. We find that stores in the general-merchandise subsector, which includes chains that sell a wide breadth of products, are not only larger than stores in other retail subsectors by every available measure, they are also growing much faster. Moreover, general-merchandise chains operate more stores than other chains, and are growing fast in this dimension as well. These trends are complementary: the same firms that are growing in one dimension are also growing in the other. Using firm-level records spanning thirty years for the general-merchandise subsector from the Census of Retail Trade, we document a strong and persistent relationship between a firm’s scale and a newly developed measure of scope in a difference-in-difference specification.

The fact that these trends are strongest among general merchandisers suggests an explanation that relies in part on a “one-stop-shopping” effect. Consumers’ desire to consolidate shopping trips is a motivator for product-line expansion outside general merchandising as well, but it is likely to be strongest in a subsector that already, by definition, sells a wide variety of products. What we have in mind for product proliferation is not the phenomenon of increased varieties within a narrow product category, such as more sizes, flavors, and brands of breakfast cereals, which [Betancourt and Gautschi \(1990\)](#) call product depth. Our focus, instead, is on increased product *breadth*, the introduction of completely new and seemingly unrelated product classes into a single store. Whereas increased product depth provides more opportunities for substitution across products sold by the same store, greater product breadth increases the total number of products a consumer buys in the store and reduces both the number of shopping trips she makes and her corresponding transportation cost.

Our empirical work uses micro data collected at the store level from the quinquennial Economic Census, starting with 1977 data and continuing through 2007. The data allow us to track general-merchandise retailers over time and follow their expansion into new product markets as well as new geographic markets, through the addition of stores. The data are uniquely suited to study the expansion of product breadth because the forms require retailers to provide information on their sales receipts from each of approximately 40 broad product categories, including groceries, women’s apparel, hardware, and optical goods. We find that firms’ expansion on the two margins — number of stores and number of product lines — are complements. Evaluating our point estimates for the average chain, adding two stores (a 10% increase in the number of stores) is associated with an addition of roughly 100 items to each store in the chain.

To explain our empirical observations, we build a simple model of a retail firm simultaneously choosing its scale (the number of stores it operates) and scope (the number of distinct product lines it carries), in which the firm’s two margins for expansion are strategic complements. Technological innovations such as bar codes and electronic data interchanges have facilitated automated purchasing from suppliers and reduced the cost of inventory management, allowing chains to take advantage of economies of scale in purchasing without undue logistical costs. We argue that because of the interaction of economies of scale on the cost side with a demand for one-stop shopping, any innovation that directly increases the number of stores a general-merchandise chain operates also induces the chain to expand the range of products it sells. On the cost side, as a chain adds stores — increasing sales volume — economies of scale cause the marginal cost of the product to fall. Lower marginal cost induces the chain to increase its range of product offerings, drawing in more customers who take advantage of its “one-stop” offering. The combination of more products and more customers at each store increases store profit, prompting the retail chain to add even more stores.

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 and abstracts from the chain's choice of product selection. [Holmes \(2011\)](#), for example, models the spatial expansion of Wal-Mart in the presence of "economies of density." Work on large or multi-product retail stores, in contrast, tends to treat stores as independent entities. Recent papers include [Holmes \(2001\)](#), who argues that improvements in inventory-management technology such as bar-code scanners have led retailers to expand product breadth. [Marvel and Peck \(2008\)](#), in contrast, show that under some conditions these same technology improvements can have the opposite effect, reducing product depth for many retailers. [Ellickson \(2007\)](#), who focuses on the supermarket business, treats expanding variety as an investment in quality that is an endogenous fixed cost which creates natural oligopolies. An important exception is [Betancourt \(2005, pp. 173-180\)](#), who explicitly models a chain's choice of scope and notes the symmetries between scope and scale, but does not discuss the complementarities between them. By formulating the retail chain's problem as a simultaneous choice of scale and scope, we can analyze both the direct effect of technology on scale and scope and the interaction between scale and scope, and bring together previously disconnected strands of the literature on the effects of chain stores.

The rest of the paper is organized as follows. Section 2 provides an overview of the general-merchandise subsector and presents the aggregate stylized facts motivating our model. Section 3 describes the Census micro data and shows the relationship within general merchandising between chain growth and product-line growth. We lay out our theoretical model of chain retailing and competition and derive conditions for complementarity of a dominant retailer's store size (scope) and its chain size (scale) in Section 4, and offer some extensions in Section 5. Section 6 concludes with a discussion of the implications of our results.

2 Background and Aggregate Stylized Facts

General merchandisers have become increasingly important in retail, accounting for nearly 15% of all retail revenues in 2007, including 75% of children’s wear, 72% of toys and hobby goods, 51% of paper products, and 49% of small electrical appliances.¹ A general-merchandise store is also more likely than a store outside of general merchandising to belong to a chain. Table 1 provides a snapshot of the differences between the retail sector as a whole and the general-merchandise subsector from the 2007 Census of Retail Trade (CRT).² General-merchandise retailers constitute only 1.2% of all retail firms, but account for 4.1% of retail stores and more than three times as much in terms of revenue. In 2007, 81% of general-merchandise stores belonged to chains but only 39% of stores outside general merchandising belonged to chains; the corresponding figures for chains with at least 100 stores were 77% and 24%.

In this section, we establish three stylized facts using the aggregate data. First, the retail sector as a whole has become more concentrated over time. Second, this trend is particularly striking among general merchandisers. Third, stores have also grown over time, increasing their square footage, employment, sales, and product breadth; this has been particularly true for general-merchandise stores. We discuss each of these facts in turn.

First, the retail sector as a whole has become much more concentrated over time, with retail chains accounting for a growing share of consumer dollars. Figure 1(a) documents the rising share of retail revenues accounted for by retail chains. Until the late 1970s, more than

¹The North American Industrial Classification System (NAICS), currently used by the Census, defines “General Merchandise Stores” (NAICS 452) 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 staff capable of retailing a large variety of goods from a single location” ([Office of Management and Budget, 1998](#), p. 445). Before the classification change that took place in 1997, the Census used the Standard Industrial Classification (SIC) system, which defined the subsector (SIC 53) 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](#), p. 317).

²Unless otherwise specified, all figures and tables in this section were created from published CRT tables and not directly from the firm-level data we use later in this paper. We rely on published Census of Retail Trade figures whenever possible to comply with Census Bureau policy.

half of 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, with 100 or more stores, more than tripled over the same period. Along similar lines, [Jarmin, Klimek, and Miranda \(2009\)](#) show, using data from the Census Bureau's Longitudinal Business Database, that chains' share of the total retail store count and their share of total retail employment has been increasing for several decades.³

Empirical studies of the competitive effects of large general-merchandise chains such as Wal-Mart and Kmart have focused on easy-to-measure outcomes. Several studies analyze store closings in the immediate vicinity of the new superstores. [Basker \(2005\)](#) estimates that each new Wal-Mart store accounts for, on average, a net reduction of 4.7 stores with fewer than 100 employees. Over the course of the thirty years from 1977 to 2007, Wal-Mart opened approximately 3,000 new stores ([Basker and Noel, 2009](#)), which, by this estimate, have caused the closure of over 12,000 competing stores. On a similar scale, [Jia \(2008\)](#) finds that Wal-Mart's expansion between 1987 and 1997 explains approximately half of the decline in single-store general merchandisers in the U.S. over this time period, and 30–40% of the decline of smaller general-merchandise chains. 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 2007.

The second stylized fact to emerge from the aggregate data is that general-merchandise chains have been growing much faster than specialist chains.⁴ 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: 99% of all general-merchandise dollars are spent at chains, and 97% at large chains, again defined

³For a historical perspective on the difficulties of building up a general-merchandise retail chain, see [Chandler \(1969\)](#); [Raff and Temin \(1999\)](#).

⁴We use the term “specialists” to include any subsector in retail other than general merchandisers and non-store retailers such as catalog companies, retail web sites, and vending machine operators.

as chains with 100 or more stores, up from 34% fifty years ago. Figure 2(a) shows the average number of stores operated by general-merchandise firms and all other retailers. Between 1963 and 2007 the average number of stores in a general-merchandise retailer more than tripled, rising from 1.43 to 5.15; average chain size also increased in the rest of the retail sector, but only modestly, from 1.12 to 1.56.⁵

As Table 2 documents, over the course of the thirty years from 1977 to 2007, the number of single-store general merchandisers has fallen by more than 40% and the number of small chains by over 75% while the number of large chains has decreased by 20%. This consolidation is not a symptom of a shrinking subsector: revenues in real terms have increased by 80%. CRT data show that in 2007, the top four general-merchandise chains alone accounted for more than 10% of all retail revenue, more than double their share thirty years earlier.

The aggregate data also show our third stylized fact: the average retail store has grown, and the average general-merchandise store has grown even more. This is true across a number of measures of store size: employment, revenue, product selection, and square footage. [Jarmin, Klimek, and Miranda \(2009\)](#) document the growth in store-level employment using data from 1958 to 2000. On average, employment per store has more than doubled, from six to 14, over this time period; the biggest increases have occurred at national chains. Although part of the explanation for this trend could be that stores are open longer hours, [Jarmin, Klimek, and Miranda \(2009\)](#) also show that, at the county level, the average number of stores per capita has been decreasing, consistent with increased consolidation rather than just an increase in hours of operation.

Revenues, another measure of store size, have also been increasing throughout the retail sector, but most noticeably at general-merchandise stores. Figure 2(b) shows the trends in revenue per store in general merchandising, food stores, and the rest of retail from 1954

⁵We exclude eating and drinking establishments and non-store retailers from the “rest of retail.” Eating and drinking establishments were classified as retail establishments under the SIC, but are no longer included as retailers in NAICS.

to 2007, in millions of 2007 dollars (CPI deflated). While revenue per store in other retail subsectors have increased by 1.3 log points, nearly quadrupling, in real terms over this period, general-merchandisers' revenue per store, already twice as high than other retailers' in 1954, have increased by nearly two log points (six-fold) over the same time period, increasing from \$2m in 1954 to \$12.6m in 2007.

Among general merchandisers, there is also evidence of square-footage growth and growth in product breadth. The most obvious measure of size is store square footage. In 1977, 53.7% of general-merchandise stores had total floor space exceeding 50,000 square feet; by 2007, that fraction was up to 61.5% of general-merchandise stores.⁶ Larger stores carry more distinct product lines: stores larger than 50,000 square feet carry 25–70% more distinct product lines than smaller stores. Using Census data to track product selection over time is harder, because changes in Census forms and in the definition of product lines affect the observed probability that a store sells any given product. For this reason, when we turn to the micro data in the next section, we control for changes in forms over time. However, there is both systematic and anecdotal evidence that the number of products carried by general merchandisers has increased over time. Figure 3 shows the fraction of general-merchandise stores selling selected lines in 1977, 1992, and 2007. The share of general-merchandise stores selling groceries increased from 64% to 88% between 1977 and 2007, and the share selling furniture increased from 42% to 55%. Anecdotally, moreover, many general merchandisers have added groceries to their product offerings; since the late 1980s, Wal-Mart, Target, and Kmart have all added “superstore” formats that include full grocery stores ([Basker and Noel, 2009](#)).

From the above discussion, we distill the following: that general-merchandise chains, which by definition sell a wide variety of products, tend to be larger, and have been growing faster, than other retail chains, and that stores, in particular within general merchandise,

⁶These numbers are calculated from micro data files, using weights to account for sampling. We do not have reliable micro data on square footage outside general merchandising.

have been growing. To see whether these trends — the growth of retail chains and the rise of the “big-box” retail format — are occurring in the same chains, we turn in the next section to Census of Retail Trade micro data. Using data on product sales and the number of stores for general-merchandise firms in the United States over a 30-year period, we show that there is a systematic relationship between growth in firm size and growth in the number of distinct product categories sold by the firm.

3 Empirical Analysis of Census Micro Data

3.1 Census Micro Data

In this section, we use micro data from the Census of Retail Trade (CRT) for the years 1977, 1982, 1987, 1992, 1997, 2002, and 2007.⁷ Detailed Economic Census forms, from which our data are compiled, are mailed to every chain store, as well as to a sample of single-store retailers. Data on stores that do not receive or do not return forms comes from administrative records. To comply with Census Bureau confidentiality restrictions we cannot report any information that can be linked to any specific firm, nor can we provide a list of firms included in the analysis.

The unit of observation in the CRT is the establishment, or store. We limit our analysis to stores in general merchandising, SIC 53 (or NAICS 452) that received Census forms numbered RT-5301 or RT-5302 (in 1982–1997) or RT-45201 or RT-45202 (in 2002 and 2007).⁸ We are able to aggregate store-level information to the firm level by using firm identifiers for each store. Firm identifiers allow us to track ownership over time, so we know the size, or scale, of each chain (number of stores it operates).

A second 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

⁷See [Foster, Haltiwanger, and Krizan \(2006\)](#) for more details on the CRT.

⁸We do not restrict observations by form in 1977 because form information is not available for that year.

“toys, hobby goods, and games” and “major household appliances.” There are two types of product lines: **broad lines** and **detailed lines**, which provide a partial breakdown of the broad lines. Table B-1 in the Appendix 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.⁹ 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 and 2007, however, and only for stores receiving form RT-45202; all other forms include the broad line “groceries” without the detailed breakdown. We drop from our dataset a small number of stores that report selling just one product line on the grounds that they are either misreporting or incorrectly classified as general merchandisers.

Taking a simple count of the number of lines over time is problematic, because the number of lines listed on the general-merchandising forms changes over time as well as across the two forms at a point in time. We take several steps to adjust for increases over time in the number of lines listed on the Census forms. To minimize the increase in reported product lines, we created a longitudinally consistent concordance of product lines. In addition, we control for year×fraction of establishments in each firm receiving each form in our regression analysis to capture any changes in the forms over time.

In most cases, firms report product-line revenues for all their stores. However, we do control for a firm’s **coverage**: the share of its stores reporting lines data. More than 95% of retailers in our dataset have coverage of 1, and more than 99% have coverage above 0.5 (we omit retailers that do not report lines data for any of their stores from this calculation). An implicit assumption is that the reporting stores are representative of the chain as a whole. We impute lines only in a few special circumstances. First, if a store reports sales of a broad

⁹Product data have never been used before in a micro-data study of the retail sector. [Basker and Van \(2010\)](#) use aggregate information on product-line revenues by different retail subsectors. Product-level information has been used in many studies of manufacturing firms; see [Foster, Haltiwanger, and Syverson \(2008\)](#), for example. [Yeap \(2011\)](#) uses product-level information in a study of the service industry.

product category but does not indicate which detailed line(s) it sells, we add one to its count of detailed lines. We also add a detailed line to the count if the revenue shares of all detailed lines within a broad category sum to less than 98% of the revenue share reported for the broad line. We add a broad line if a store reports selling a detailed product line but not its parent broad line. Finally, if a store reports selling one or more product lines that do not appear on the form it received (“write ins”) we add one detailed line and one broad line to its count of lines.

The count of lines sold at the firm level is an upper bound on the actual number of lines sold. Some lines may be experimental, appearing in a small fraction of stores and generating little revenue, or they may be reported in error by one or two stores in a chain. This type of “line inflation” is a particular concern for the current analysis because it is likely to be more pronounced for larger chains, where more stores mean more forms with the opportunity for error. To overcome this problem we calculate a weighted average of lines, weighted by the fraction of stores in the chain selling that line. This number is identical to the total number of lines sold for single-store retailers and for retailers selling the same lines in each store, but it falls below the total number of lines for chains selling either a different number of lines or a different composition of lines across stores.

Finally, as a robustness check, we also calculate the “effective” number of lines sold by each firm using the inverse of the product-level Herfindahl index at the firm level:

$$\text{Inverse Herfindahl} \equiv \frac{1}{\sum_{i=1}^n \left(\frac{\text{revenue}_i}{\sum_{j=1}^n \text{revenue}_j} \right)^2}$$

where n is the total number of lines sold by the retailer, and revenue_i is the revenue the firm receives from sales of product i . The inverse Herfindahl takes on values in $(1, n]$. The upper bound is obtained only in the case of equal revenue shares across all products. The more

sales are dominated by a single product, the closer is the inverse Herfindahl to one.¹⁰

Table 3 provides summary statistics of the data used in the regressions. The data set represents a snapshot of the general-merchandise subsector in each of the seven census years. Of the nearly 17,500 observations in our data, representing nearly 13,000 retailers, more than 14,000 are single-store retailer-years, and almost 3,000 retailer-years are small chains, averaging 8.9 stores per chain. Only 235 observations come from chains with 100 or more stores, but these chains have 628 stores on average. Both the number of broad lines and the number of detailed lines increase with chain size. As expected, the total number of products sold is larger than the average number of products across stores in the firm, except in the case of single-store retailers, where they are identical. Also as expected, the average number of lines sold is less sensitive to chain size than the total number of lines sold.

Due to entry, exit, and consolidations, our data comprise an unbalanced panel: some firms show up only once, while others appear in all seven census years. The last column in Table 3 provides the share of firms (by size class) that are **dynamic**, that is, for which we have more than one observation in the data. We include this information to clarify which observations are contributing to the identification of the parameter of interest in the regression analysis in the next section. Virtually all dynamic firms exhibit changes in the number of stores and the number of lines over the study period and therefore contribute to the estimation of the relationship between firm scale and scope in our regressions, which always include firm fixed effects. Almost all large chains (93%) are dynamic, as are 73% of small chains and 38% of single-store retailers. The second panel of Table 3 provides a closer

¹⁰Many combinations of revenue shares produce the same Herfindahl value. For example, consider firms selling 6.4 effective broad lines (the average number for dynamic firms in our dataset). All of them must sell at least seven products in total, but depending on the distribution of revenue shares, the total number of products could vary considerably. A firm selling seven products, one with revenue share of 25%, and each of the remaining six with revenue share of 12.5%, has 6.4 effective product lines; so does a firm selling 14 products, one with revenue share of 35% and each of the remaining 13 with revenue share of 5%. The inverse Herfindahl may, hypothetically, even fall when a new line is introduced by a retailer, if the new line turns out to dominate sales. In practice, the number of effective lines is highly and positively correlated with the total, and average, number of lines sold by the retailer.

look at the dynamic subsample of the data. This sample contains about 7,700 observations from 3,000 firms, which are, on average, larger than the firms in the full sample and carry more lines.

The total number of observations in Table 3 is smaller than the one reported in Table 2 for two reasons. First, as noted above, we omit firms that did not get the general-merchandise forms from our data set. Second, we exclude retailers for which no store reported any product line sales. This is much more likely to happen for single-store retailers, of which we use approximately 20%. In contrast, virtually all the large-chain observations from the CRT are included in our data.

3.2 The Relationship between Store Growth and Line Growth

We start with a simple descriptive analysis, testing for a correlation between a retailer’s log number of stores and the log of the number of product lines it carries: $\rho(\ln(\text{stores}), \ln(\text{lines}))$. The correlations are all positive and statistically significant at the 0.01% level, and range from 0.20 (for both the average and “effective” number of broad lines) to 0.31 (for the total number of detailed lines). We repeat the same exercise for log changes, calculating $\rho(\Delta \ln(\text{stores}), \Delta \ln(\text{lines}))$. Again, we find very significant, albeit smaller, correlations between the change in the log number of stores and the change in the log number of product lines carried, with correlation coefficients ranging from 0.05 (for the average number of detailed lines) to 0.12 (for the total number of both broad and detailed lines).

Next, we compare the probability of an increase in the number of product lines with and without a concurrent increase in the number of stores. To do this, we separate retailers into those whose store count grew and those whose store count stagnated or decreased since the previous census year. Let $\mathbf{1}(\Delta \text{stores}_{rt} > 0)$ be an indicator for retailer r ’s store count increasing between year $(t - 5)$ and year t , and define $\mathbf{1}(\Delta \text{lines}_{rt} > 0)$ similarly, for the number of lines. Forty-three percent of continuing firms that did not add stores increased their line count, compared to 51% of stores that did add stores. This difference is statistically

significant at the 1% level. For detailed lines, the difference is smaller (0.52 vs. 0.57) but still statistically significant at the 5% level.

We use a difference-in-difference specification to put the analysis into a regression framework. This allows us to estimate the magnitude of the relationship between stores and lines while controlling 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 sold and number of stores operated. We estimate

$$\ln(\mathbf{lines}_{rt}) = \alpha_r + \sum_f (\theta_t \cdot \phi_{rft}) + \beta \cdot \ln(\mathbf{stores}_{rt}) + \gamma \cdot \mathbf{coverage}_{rt} + \varepsilon_{rt} \quad (1)$$

where \mathbf{lines}_{rt} is the number of lines of merchandise sold by retailer r in year t , α_r is a retailer fixed effect for each of the 12,787 retailers in the sample, θ_t is a year fixed effect for each of the seven census years, ϕ_{rft} the fraction of stores belonging to retailer r that received form f in year t (for forms RT-5301, RT-5302, RT-45201, and RT-45202), and \mathbf{stores}_{rt} is the number of stores operated by retailer r in year t . The retailer fixed effects ensure that we identify the relationship between stores and lines using only within-firm variation over time; only dynamic firms contribute to the parameter identification. This means that although the regression is in levels, we can interpret β as the relationship between *changes* in $\ln(\mathbf{stores})$ and *changes* in $\ln(\mathbf{lines})$. 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 a retailer's stores for which valid line counts are obtained. The error term ε_{rt} is clustered at the firm level to allow for arbitrary autocorrelation.

Table 4 reports our estimates of the coefficient β for the full sample. Each column represents a different measure of product lines. Results using our preferred measure of product lines, the average number of lines sold across stores in the chain, are presented in the first two columns: first broad lines, then detailed lines. The elasticity estimates in both

cases are near 0.07. In other words, a 10% increase in the number of stores is associated with a 0.7% increase in the number of product lines.

Recall that the average dynamic firm in our data in Table 3 has 22 stores, so a 10% increase in its store count amounts to two new stores. Applying the average line counts for dynamic firms, our estimates imply that this two-store increase is associated with, respectively, a 0.1 increase in the average number of broad lines and a 0.15 increase in the average number of detailed lines sold by stores in the chain.

How many items does a tenth of a broad product line represent? Census data do not break down detailed lines into stock-keeping units (SKUs), but we have some sense of this from outside sources. A typical Wal-Mart store, for example, sells as many as 1,400 book titles (Heher, 2009) and 1,800 different toys (Reuters, 2002).¹¹ A Target store is said to carry between 600 and 1,500 music SKUs (Christman, 2002). Other product categories may be significantly smaller (e.g., large appliances) or larger: the number of grocery products in a typical “Supercenter” is well above 10,000. Assuming, for convenience, that a typical product category has 1,000 products, each 0.1 increase in a broad product line amounts to selling 100 additional items. Two new stores in a chain of 22, then, are associated with an average increase of approximately 100 items per store in the chain.

The next set of estimates allows us to determine whether this increase reflects the same 100 or so products being added at all stores in the chain, or 100 different products in each store. In the first case, the total number of lines sold by the chain would increase to the same degree as the average; in the second case, it would increase twenty times as much. Replacing the average number of lines sold per store with the total number of lines sold chain-wide as the left-hand side variable does increase the point estimates, but by a factor just shy of two. Part of this increase may be due to measurement error in the total line count, which, as

¹¹Forman, Ghose, and Goldfarb (2009) visited two Wal-Mart stores and two Target stores and counted the number of book titles each offered. They found about 850 titles in the Wal-Mart stores and about 1,300 titles in the Target stores.

noted earlier, is exacerbated by the size of the chain. Although it is impossible to say with certainty how much of the increase in the coefficient estimates is due to this bias and how much to product diversification across stores in the chain, the results suggest that most, if not all, of the increase represents the same product lines being added to all stores in the chain.

The last two columns use the “effective” number of product lines sold by the chain, calculated as the inverse Herfindahl index. The fact that the inverse Herfindahl increases as much as, or more than, the average line count indicates that the new products reduce the revenue share of dominant products and level the revenue distribution across products.

In addition to estimating Equation (1) on the full set of general-merchandise firms, we have also estimated it separately on subsets of firms by size. Given the dominance of single-store retailers and small chains in the firm population, it is not too surprising that we find very similar estimates when we restrict the sample to small chains. The estimates for large chains, which we show in Table 5, are somewhat larger. For the average number of product lines, the coefficients for the large-chain sample are approximately 50% larger than for the full sample. The average large dynamic chain has 660 stores and carries an average of 21 broad product lines in each store (or 28 broad lines throughout the chain). For this chain adding 66 stores, which increases the number of stores by 10%, is associated with an increase of about 1.3% in the average number of broad lines, or more than a quarter of a broad line, carried by the chain. This amounts to an increase of about 270 individual items per store. The coefficient estimates with total lines and the inverse Herfindahl as left-hand side variables are extremely similar to the results using average line counts, again consistent with the same lines being added across stores in the chain and with a leveling of the revenue distribution.

The economic repercussions of even a small increase in the number of product lines carried by a general-merchandise chain can be quite large. In the famous case of Wal-Mart’s addition of a full line of groceries to its “Supercenters,” Wal-Mart has nearly tripled its store

count over a twenty-year period while adding grocery lines to most, but not yet all, its stores (Basker and Noel, 2009), and continuing to add other lines, such as pharmacies, a process started even earlier. This process has been slow, consistent with our coefficient estimates, but has had far-reaching effects on consumers' access to groceries, on prices, selection, and quality industry-wide. Incumbent supermarkets have reduced their prices by 1–2% on average when a new Wal-Mart Supercenter opens in their city (Basker and Noel, 2009) and reduced their stockouts by up to 24% compared to pre-entry levels (Matsa, forthcoming).

As general merchandisers expand into an increasing number of product lines, subsector after subsector realigns to adjust to the new competitive environment. In the book market, Forman, Ghose, and Goldfarb (2009) find that Wal-Mart and Target entry increase relative demand at Amazon.com, and by implication at local bookstores as well, for obscure titles over best sellers. Inventory costs are likely to rise if bookstores sell fewer best sellers, and lower demand reduces revenue. Both effects reduce the profitability of incumbent bookstores.

In the next subsection, we explore the role that mergers and acquisitions play in the growth of both stores and lines. Following that, we take a detailed look at the extent to which firms are adding lines in new versus in existing stores.

3.3 Mergers and Acquisitions

Anecdotally, the largest general-merchandise chains have grown both by building new stores and by acquiring stores, and sometimes entire chains. Examples of the latter range from Wal-Mart's purchase of 92 Kuhn's-Big K stores in 1981, which increased Wal-Mart's store count by more than 25% overnight (Wal-Mart Stores, Inc., 1982), to the 2005 mega-merger of Sears and Kmart that created the third-largest retailer in the United States (Irwin, 2005).

We test whether mergers and acquisitions (M&A) represent a special opportunity to increase product lines, even after controlling for the number of stores in the chain, by adding a **M&A** indicator to the regression from Equation (1). We construct this indicator by

observing whether any store belonging to retailer r in year t belonged to a different retailer in the previous census year.

The results are shown in Table 6. We find that the effect of a merger or acquisition on the average number of lines sold by stores in the chain is as large as the effect of doubling the chain size. Whereas the total number of lines increases by more than the average number of product lines in conjunction with a given increase in store count, the effect of a merger on the two outcomes is quite similar, consistent with adding the same lines to all stores in the chain. The effect of a merger on the inverse Herfindahl measure of lines is smaller and less precisely measured. This last effect is consistent with the ambiguous relationship between the inverse Herfindahl and the number of product lines.

3.4 New Stores

The number of product lines in a growing chain can increase because new stores have more products lines than existing stores, because the number of product lines in existing stores increases, or because of a combination of these. To investigate the relative importance of new stores in increasing chains' product breadth, we decompose increases in the number of product lines into the share due to increases in product lines in continuing stores and the difference between new stores' product lines and the average number of product lines sold by the chain in the previous census year.

Formally, let \mathbf{lines}_{rt} be the number of product lines sold by retailer r in year t , averaged across stores in the firm, and $\mathbf{lines}_{rt}^{\text{cont}}$ and $\mathbf{lines}_{rt}^{\text{new}}$, respectively, be the average number of lines sold by continuing and new stores in retailer r . Finally, define,

$$\begin{aligned}\Delta\mathbf{lines}_{rt} &\equiv \mathbf{lines}_{rt} - \mathbf{lines}_{r,t-1} \\ \Delta\mathbf{lines}_{rt}^{\text{cont}} &\equiv \mathbf{lines}_{rt}^{\text{cont}} - \mathbf{lines}_{r,t-1} \\ \Delta\mathbf{lines}_{rt}^{\text{new}} &\equiv \mathbf{lines}_{rt}^{\text{new}} - \mathbf{lines}_{r,t-1}.\end{aligned}$$

Then the following decomposition holds as an accounting identity:

$$\Delta \text{lines}_{rt} \equiv \underbrace{w_{rt}^{\text{new}} \cdot \Delta \text{lines}_{rt}^{\text{new}}}_{\text{New Share}} + \underbrace{w_{rt}^{\text{cont}} \cdot \Delta \text{lines}_{rt}^{\text{cont}}}_{\text{Continuing Share}} \quad (2)$$

where the weights, w_{rt}^{new} and w_{rt}^{cont} represent the relative counts of new and continuing stores, respectively, in year t . We restrict this analysis to the case where the average number of lines is non-decreasing, for the chain as a whole as well as for its new and continuing components, to ensure that our weights sum to one.

We can identify three sorts of “new” stores in the chain. First are stores that are newly constructed, or “new-new” stores. Second are stores that were added to the chain in a merger or acquisition, as discussed in the previous section, or “merged-new” stores. Third are stores that moved a short distance or underwent a major renovation, or “new-and-improved” stores. We have already described how we identify the first two types in previous sections. For the third type, new-and-improved stores, we rely on store-level reports of square footage on Census forms. Approximately two-thirds of general-merchandise stores in our data report total under-roof square footage. When we have reports for two consecutive years, we interpret a 10% or larger change in total square footage as a major renovation or address change.¹²

We find that new-new stores contribute 31% of the growth in product lines at the retailer level. This value is invariant to the use of broad or detailed product lines. Merged-new stores make a negligible contribution despite the fact that, as found in the previous subsection, mergers themselves are correlated with relatively large increases in line counts. New-and-improved stores contribute 15% of broad-line growth and 18% of detailed-line growth. Continuing stores in continuous ownership and with no observed changes in square footage

¹²In principle, we could use address information to determine moves, but in practice this is not feasible due to the very large number of variations for individual addresses, ranging from abbreviations, including or excluding unit numbers in malls, spelling variations, etc.

contribute approximately 50% of the total.

What this means is that, as a general-merchandise retailer grows, it not only increases the product lines it sells in its new stores, but also in its existing stores. This, combined with the preceding regression analysis, suggests the presence of some economies of scope in selling individual product lines. When a retailer adds a product line, it adds the line throughout the chain, although in some cases this process may take years to complete.

The next section presents a model of a chain retailer motivated by the above empirical analysis. In equilibrium, the chain retailer’s choice of the number of products to sell and the number of stores to operate turn out to be strategic complements, and any increase in parameter values — for example cost or technology parameters — leads the chain to increase its scale and scope together.

4 Model

4.1 Environment

There is a large set of non-overlapping *ex-ante* identical locations. In each location, a continuum of consumers has unit demand for each of N product categories, or lines, such as apparel, toys, or hardware. Our goal is to capture the sorts of product categories we have in the Retail Census data set, and the fact that a shopper at a general-merchandise store is likely to buy a basket of products — perhaps a pair of shoes, a toy, a microwave oven, and a book — unlike a shopper at a shoe store.

There are two types of stores: mom-and-pop specialist stores, each of which sells one line in a single location; and a chain superstore selling multiple lines at multiple locations.¹³ Every trip to a store entails a transportation cost, which includes both the time cost and

¹³The assumption that specialist stores do not belong to chains is consistent with the evidence in Figure 2(a) that the average number of stores operated by a specialist retailer is well below two. We discuss specialist chains in an extension in Section 5.2.

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. In this paper we interpret the “transportation cost” very broadly to include any transaction cost, such as standing in line and digging for change, that needs to be repeated at every stop.¹⁴ Although we have general-merchandise stores in mind when we think about one-stop shopping, the same effect may exist to a lesser extent in other subsectors, chiefly groceries.

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, σ , is distributed according to the exponential probability density function

$$f_{\sigma}(\sigma) = \exp(-\sigma).$$

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 δ ; and $w(\cdot)$ is the variable cost of selling a given volume of a single line. The assumption that the chain sells the same n products in each of its k stores is justified by the fact that the total number of product lines sold by a chain in the Census micro data does not diverge by much from the average number of product lines sold across stores in the chain (see Table 3).

¹⁴Betancourt and Gautschi (1988, 1990) rigorously model the implications of transportation costs and other distribution services for the household’s production function, but they consider only a one-store 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 an “implicit good” by Betancourt and Gautschi) and prices. Klemperer and Padilla (1997) discuss the related idea of consumers’ “shopping costs” giving rise to a consumer’s preference to use one supplier. However, there, the shopping cost comes from establishing new relationships, uncertainty, learning about new suppliers, and feelings of loyalty, which makes it closer in spirit to the idea of a “switching cost” rather than the transportation cost in our model.

The variable cost $w(\cdot)$ is positive, increasing, and concave, with $w'(0) < c_m^i$ for all i . One reason for a decreasing marginal cost is that larger retailers may be able to lower costs by engaging in direct negotiation with alternative suppliers, especially ones located in more remote areas (e.g., in less developed countries); this idea is explored by [Basker and Van \(2008, 2010\)](#) and [Caprice and von Schlippenbach \(2008\)](#). In addition, increased concentration in the retail sector may decrease suppliers' outside options, so that in a bargaining situation they may be forced to accept a lower price ([Inderst and Wey, 2007](#); [Caprice and von Schlippenbach, 2008](#)).

The coordination cost $d(\cdot, \cdot)$ includes what [Bliss \(1988\)](#) calls “overhead cost,” as well as costs associated with agency and logistics.¹⁵ We assume that the coordination cost increases with the number of stores in the chain and with the number of product lines the chain carries, that is, $\frac{\partial d}{\partial n} > 0$, $\frac{\partial d}{\partial k} > 0$. In general, the cross-partial derivative, $\frac{\partial^2 d(n,k)}{\partial k \partial n}$, can be positive, negative or zero. If $\frac{\partial^2 d(n,k)}{\partial k \partial n} < 0$, increasing the number of product lines (stores) reduces the added cost of adding a store (product line). That is, there is a positive spillover between adding lines and stores in terms of coordination. If $\frac{\partial^2 d(n,k)}{\partial k \partial n} > 0$, adding product lines (stores) has a negative spillover on adding stores (product lines). We are agnostic on whether any such spillovers exist and, if they exist, whether they are negative or positive. Our results require, however, that any negative spillovers in the coordination of lines and chains not be too big, at least for the relevant ranges of n and k . In other words, we assume that $\frac{\partial^2 d(n,k)}{\partial k \partial n} \leq R$, where R is a positive constant.¹⁶ We make no assumptions about the second derivatives $\frac{\partial^2 d(n,k)}{\partial k^2}$ and $\frac{\partial^2 d(n,k)}{\partial n^2}$. The parameter δ indexes the firm's technology; the larger the value of δ , the more advanced the firm's technology with respect to inventory management, logistics, and distribution, and the lower is the cost of adding additional stores and/or product lines.

¹⁵For a discussion of agency costs see [Lafontaine \(1992\)](#).

¹⁶Convexity of $d(\cdot, \cdot)$ would be consistent with the findings of [Kosova and Lafontaine \(2010\)](#) for franchised chains. There is a consensus in the marketing literature that adding product lines is associated with increased costs; see, for example, [Dukes, Geylani, and Srinivasan \(2009\)](#).

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 sells. In the second stage the superstore and all mom-and-pop stores simultaneously set prices. Finally, 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 4.

4.2 Consumers

Consumers take the presence of a superstore, the lines it sells, 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 cost, τ , and aversion cost, σ) of buying all N goods. Consumers shop at 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,” 22 percent named broad selection/variety ([Pew, 2005](#)).¹⁷

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 this condition is guaranteed below.) Given this condition, a consumer who

¹⁷We view malls as an imperfect substitute for superstores because, although consumers drive to one location and park there, they engage in many separate transactions. Consistent with the idea that shopping at a superstore saves time relative to the alternatives, [Carden and Courtemanche \(2009\)](#) find that Wal-Mart’s presence may increase consumption of “cultural” activities such as attending classical music concerts and visiting art galleries.

makes a trip to the superstore will buy all n items offered by the superstore for the bundle price of

$$P_s \equiv \sum_{i=1}^n p_s^i. \quad (3)$$

A consumer chooses to shop at the superstore over purchasing all N lines at mom-and-pop stores if and only if

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

which defines a threshold aversion value:

$$\bar{\sigma} \equiv (n - 1)\tau + \sum_{i=1}^n p_m^i - P_s. \quad (4)$$

Consumers with superstore 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

$$P_s \leq (n - 1)\tau + \sum_{i=1}^n p_m^i \quad (5)$$

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.

4.3 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 in equilibrium. Demand for each of the superstore's products is $x_s^i = F_\sigma(\bar{\sigma})$ while each mom-and-pop store competing with the superstore has demand $x_m^i = 1 - F_\sigma(\bar{\sigma})$. (We ignore mom-and-pop stores that do not compete with the superstore.)

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

$$p_m^{i*}(P_s) = c_m^i + 1. \tag{6}$$

In other words, the dominant strategy for mom-and-pop stores that compete directly with a superstore calls for pricing at $c_m^i + 1$. 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 functions would depend not only on the superstore's prices but also on 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 at a representative store, taking all mom-and-pop stores' prices (and its own prior choices of

scale and scope) as given. 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. Using the inverse demand function $P = P(x)$, we can simplify the superstore's optimization problem to:

$$\max_x \quad \pi = x \cdot P(x) - \frac{nw(kx)}{k} \quad (7)$$

The first-order condition for problem (7) is

$$\begin{aligned} \pi_x(x^*; p_m^{1*}, \dots, p_m^{n*}) &\equiv \frac{\partial \pi}{\partial x} \\ &= \ln(1 - x^*) + (n - 1)\tau + \sum_{i=1}^n p_m^i - \frac{x^*}{1 - x^*} - nw'(kx^*) = 0. \end{aligned} \quad (8)$$

Equation (8) implicitly 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, π^* , for given first-stage choices n and k .

The properties of the Nash equilibrium are summarized in Figure 5 depicting the two best-response functions. The mom-and-pop price, p_m^i , is on the x -axis, and the quantity sold by the superstore is on the y -axis. An 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 increase in the number of products sold, n , also draws more shoppers because of the one-stop-shopping effect.

This result guarantees that in the second stage each superstore has positive producer surplus. The chain sets n and k in the first stage, balancing the marginal benefit of each — higher operating profits — against the marginal cost associated with the increased complexity of its logistics, distribution, and management problem.

4.4 Superstore’s Scale and Scope

In the first stage the superstore chooses the number of product lines it sells, n , and the number of stores it operates, k . We assume that this decision is simultaneous, but the solution is identical if the choice is sequential (in either order). 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 the superstore’s technology is sufficiently advanced, the superstore’s scale and scope are complements. To state more precisely what we mean by sufficiently advanced technology, define

$$\bar{\delta} \equiv \frac{R}{x^*(1, 1) \cdot (\tau + c_m + 1) - w(x^*(1, 1))}, \quad (9)$$

where R is the upper bound on $\frac{\partial^2 d(n, k)}{\partial n \partial k}$.

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

To understand the mechanism behind this result, consider one condition for the supermodularity of the firm’s profit function, that the cross-partial derivative of profit with respect

to n and k be positive:

$$\frac{\partial^2 \Pi}{\partial k \partial n} = \frac{\partial \pi_s^*}{\partial n} + kx^* \cdot \left(\frac{w(kx^*)}{kx^*} - w'(kx^*) \right) - nk \frac{\partial x^*}{\partial n} w''(kx^*) - \frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k} > 0. \quad (10)$$

The first three terms contribute to this cross-partial being positively valued. The first term, $\frac{\partial \pi_s^*}{\partial n}$, captures part of the one-stop-shopping effect. When a firm increases its product offerings, the associated producer surplus increase derives from the bigger bundle now sold to existing superstore shoppers. The second term, $kx^* \left(\frac{w(kx^*)}{kx^*} - w'(kx^*) \right)$, is positive because of the economies of scale in the superstore's variable cost. The third term, $\left(-nk \frac{\partial x^*}{\partial n} w''(kx^*) \right)$, is an interaction between the one-stop-shopping effect ($\frac{\partial x^*}{\partial n} > 0$) and economies of scale in the superstore's variable cost ($w''(kx^*) < 0$).

One-stop shopping and economies of scale and the interaction between them contribute to the supermodularity of the superstore's profit function and thus the complementarity between n and k . As the chain adds more stores, its sales volume increases for each product, reducing marginal cost and increasing the total surplus it earns from its sales. This higher surplus induces the chain to add more products. Conversely, as the chain adds more products the one-stop-shopping effect shifts out demand at each of its stores, so the surplus earned at each store increases and creates an incentive for the chain to add more stores.

The fourth term, $\left(-\frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k} \right)$, is positive if $\frac{\partial^2 d(n, k)}{\partial n \partial k} < 0$, working to amplify the other three effects. When $\frac{\partial^2 d(n, k)}{\partial n \partial k} > 0$, this term operates to *dampen* the complementarity between scale and scope. A sufficient condition for a positive net effect is that $\delta > \bar{\delta}$, in which case the superstore's cost of selling a new product in a new location is lower than the cost to the consumer of acquiring the item from a mom-and-pop store, resulting in a surplus that the superstore can capture by adding another product line to one of its stores.

This result accords with anecdotal evidence about firm expansion strategies. Industry reports indicate, for example, that Wal-Mart's primary motivation for introducing the "Supercenter" format, which added a full line of groceries to the existing stores, was to increase

store traffic ([Singh, Hansen, and Blattberg, 2006](#)).

Our contribution in this model is the highlighting of the complementarity between the scope of a retailer’s operation, which provides one-stop-shopping opportunities to consumers, and its scale. Technology alone is not enough to explain the disproportionate growth of general-merchandise chains; as technology improves, complementarity of lines and stores means that the effects on these two dimensions are amplified. This effect can be seen most clearly in the higher level and disproportionate increase in chain size among general-merchandise chains as compared with other retail chains in Figure 2(a), and in the level and increase in store size and store-level revenue documented in Figure 2(b) and in the discussion in Section 2, despite having access to the same general retail technology.

5 Extensions

5.1 Endogenizing the Chain’s Technology

So far, we have treated the retailer’s technology as exogenous, which may be justified in a model in which firms get exogenous technology draws, à la [Jovanovic \(1982\)](#). Alternatively, [Bagwell, Ramey, and Spulber \(1997\)](#) model retail firms that endogenously make investments in cost-reducing technology, in which the equilibrium is characterized by mixed strategies. The investment level of any given firm may, in that case, be taken as a random draw from the mixed strategy space.

Result 1 continues to hold if the chain chooses its technology level, δ , either prior to or concurrent with choosing its scale and scope, trading off the benefits of better technology — namely, lower logistics and distribution costs — against the cost of adopting expensive technologies. Treating technology as endogenous implies that retailers in different subsectors display different levels of technology investment, corresponding to their benefits from technology adoption. (This idea is similar to the idea of “directed technical change” in recent models of endogenous technological change; see [Acemoglu 2002](#).) The bigger is the

firm's network of stores and suppliers, the greater is its incentive to invest in technology that reduces the cost of operating this network.

Consistent with this broader interpretation of technology, our empirical results show a stronger relationship between store count and line count for the largest general-merchandise firms (Table 5), which have had the most reason to make larger investments in technology. More generally, the general-merchandise subsector has been a leader in technological innovation. Evidence on technology adoption rates by retail subsector is limited, but the existing evidence points strongly to general merchandisers as leaders in this regard as well. Using data for 1992, [Doms, Jarmin, and Klimek \(2004\)](#) find that general merchandisers invest more in technology than other retailers. Evidence from the Census of Retail Trade also confirms that labor productivity growth in general merchandising has far outpaced the rest of the retail sector ([Basker, 2007](#)).

Bar-code scanners are among the biggest technological innovations affecting the retail industry in the last half century. The original motivation for this technological innovation, first implemented in the mid-1970s, was reducing cashier errors and increasing check-out speed ([Levin, Levin, and Meisel, 1987](#); [Das, Falaris, and Mulligan, 2009](#); [Basker, 2011](#)). In the early 1980s, mass merchandisers began to realize the full inventory-management potential of scanners, and started requiring suppliers to print bar codes on products; grocery stores soon followed their lead ([Abernathy, Dunlop, Hammond, and Weil, 1999](#)). Kmart and Wal-Mart, in particular, are widely credited with implementing the adoption of bar codes beyond the grocery product line ([Abernathy, Dunlop, Hammond, and Weil, 1999](#); [Dunlop and Rivkin, 1997](#)); [Holmes \(2001\)](#) uses this observation to motivate his theoretical treatment of the effect of barcode technology on product breadth.¹⁸ These general merchandisers had a strong interest in streamlining their inventory systems, which carried at the time as many as 10,000

¹⁸Interestingly, [Beck, Grajek, and Wey \(2011\)](#), who study the international diffusion of scanning technology, find that bar-code scanners diffused more slowly in countries with “hypermarket” (superstore) presence, possibly because potential adopters of the technology outside general merchandise experienced higher exit rates in response to hypermarket competition.

stock-keeping units (SKUs), compared to approximately 100,000 at a supermarket ([Harmon and Adams, 1984](#)). Today, general merchandisers are leading the push to implement the newest inventory-management innovation, radio-frequency identification tags.

Differential rates of technology adoption are one reason for the increasing size disparity between general merchandisers and other retailers. Larger (and expanding) chains throughout the retail sector are known to be more efficient than retailers operating a single store ([Foster, Haltiwanger, and Krizan, 2006](#)). Specifically, larger chains were early adopters of bar-code scanners in the 1970s and early '80s ([Levin, Levin, and Meisel, 1987](#)), and more recently they have been leaders in the implementation and use of transactional web sites, internet procurement, and data warehousing ([Hunter, 2003](#)).

Although the Census does not ask specific questions about technology adoption in the CRT, the broader Annual Capital Expenditure Survey (ACES), which has been conducted by the Census every year since 1994, asks approximately 60,000 firms their levels of equipment investment. While all firms may be included in the survey, firms with at least 500 employees are included in the sample with certainty, and ACES industry category codes identify firms whose primary activity is general-merchandise retailing. In 1998, 2003, and 2008, the forms split equipment expenditures into many detailed categories, including computers and peripheral equipment, other information and communication technology equipment, and capitalized software (including internally developed software). A supplement to ACES has captured non-capitalized hardware and software purchases since 2003. In future work, these data could provide a portal into the technology-investment decisions of large retailers.

5.2 Specialist Chains

In reality, of course, not all chains are “superstores,” although, as documented in Figure 2(a), the average retailer outside general merchandising operates at a substantially lower scale. We do not pretend to have a complete model of retailing, but in the context of our model it

is easy to see how a chain constrained to carry a single product ($n \equiv 1$) will also choose to operate fewer stores.

Although the distinction between general merchandisers and specialists in our data is based on industrial classification, as explained in Section 2, and is therefore somewhat arbitrary, it does a good job of describing stores in the specialist classifications. Evidence that stores' industrial classification is a good descriptor of their type of business is provided by [Haltiwanger, Jarmin, and Krizan \(2010\)](#), who find that large specialist chains have a negative impact on only those single-unit retailers and smaller chain stores that are classified in the same detailed subsector. Further evidence comes from the close match between a store's primary product and its industrial classification. In the 2007 CRT, for example, 68% of food and beverage stores' (NAICS 445) revenue came from food sales; 72% of book stores' (NAICS 451211) revenue came from book sales; and 88% of shoe stores' (NAICS 4482) revenue came from shoes.

The concentration of sales in a store's primary product was even higher historically. In the 1977 CRT, 86% of food stores' (SIC 54) revenue came from the sale of groceries; 85% of book stores' (SIC 5942) revenue came from book sales; and 95% of shoe stores' (SIC 566) revenue came from shoe sales. The fact that specialist stores have diversified their product sales suggests that although general merchandising is the natural place to look for one-stop-shopping effects, specialist retail chains are also trying to exploit this margin.

A thorough study of the de-specialization of specialists is beyond the scope of our paper, but we speculate that it is driven, at least in part, by a desire to exploit the same complementarities we study here in the context of general merchandisers. Anecdotally, *Supermarket News* reported in April 2009 that the toy seller Toys “Я” Us was experimenting with a new format selling “beverages, snacks, cereals, confectionery and household products” ([Supermarket News, 2009](#)). Toys “Я” Us had previously started combining its toy stores and baby stores (Babies “Я” Us) under one roof ([Rosenbloom, 2009](#)). The coffee chain Starbucks markets music CDs; the New York Times reported in 2004 that Starbucks accounted for more

than a quarter of the sales of a Ray Charles album that had just gone platinum (Levine, 2004). These chains are not alone. The 2007 CRT reports that over 12% of grocery stores sell toys, hobby goods, and games; 8% of building-materials stores sell groceries; and some enterprising fruit and vegetable markets sell jewelry. As specialist chains branch out into additional products, increasing their scope, we expect them to increase their scale as well.

5.3 Population Growth

Population growth in each location, holding the distribution of preferences constant, increases the measure of consumers who shop at the superstore in proportion to population growth and with it, the profitability of each of the chain's product lines in each of its stores, leading to an increase in both scale (k) and scope (n) (and technology, if its level is endogenous). In a different context, Campbell and Hubbard (2009) show that the size of gasoline stations increased in anticipation of increased traffic in the early years of the Interstate Highway System.

Empirically, the Census data we developed for this paper could be used to study whether retail chains with a disproportionate number of stores in growing areas add lines and stores faster than chains located in stagnant areas. However, we caution against interpreting any results causally, because chains locating in growing areas may be different in other unobserved ways from chains located in stagnant areas.

6 Concluding Remarks

Our paper makes two contributions. First, we document empirically the simultaneous increase in product breadth and chain size for general-merchandise retailers. We argue that these trends cannot be due to technology alone, because we see much weaker trends in other retail subsectors that have access to the same technology, such as the grocery subsector. We use micro data from the Census of Retail Trade to show that general merchandisers add simi-

lar lines throughout their chains when their store count increases, both in the case of mergers and acquisitions and in the case of new store construction. Second, we offer an explanation for these facts that relies on an interaction of economies of scale and demand-driven gains from scope due to consumer preference for one-stop shopping. This one-stop-shopping effect is present for other retail forms, but is strongest in general merchandising. Technological innovations, such as the introduction of Radio Frequency Identification technology in the distribution process, or cost reductions due to trade liberalization, increase both a chain's optimal scale and its optimal scope, but the interaction between the two is what sets general merchandisers apart. The interaction between the demand side and the cost side of the retailer's optimization problem 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 small businesses in an increasing number of product lines and locations. Our model implies that much of the increased competitive pressure on small retailers is due to the fact that growing chains face decreasing marginal cost curves. In addition, new product lines increase the one-stop-shopping benefits at the superstore, with negative effects on mom-and-pop stores: 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. The welfare effects of these changes can be substantial. Many of the empirical studies on the competitive impact of "big-box" stores cited in Sections 2 and 3 analyze the effect of a single general merchandiser, in some cases studying the impact of a single product line sold by a single general merchandiser. The effect aggregated over many products and locations is obviously even larger.

Other welfare implications of technological innovations and chains' resulting expansion and consolidation are more complex due to the presence of externalities. In our model, the chain's growth has only positive effects on consumers. Because a consumer choosing whether or not to shop at the superstore compares the private cost of doing so (aversion cost)

to her private benefit (lower transportation costs and prices) and ignores the externality she creates by increasing the store's sales volume (lower marginal cost of goods, which leads to lower prices on existing goods and induces the chain to add stores and products, benefiting consumers in the same location as well as across locations), fewer consumers shop at the superstore than is socially optimal. This result, however, depends critically on the assumptions we have made on the cost structure of mom-and-pop stores; specifically, the constant marginal cost and lack of fixed costs. If marginal cost for all stores were decreasing, consumers switching from shopping at mom-and-pop stores to shopping at the superstore would impose negative externalities on other mom-and-pop shoppers, offsetting positive externalities on superstore shoppers. Alternatively, if mom-and-pop stores had fixed costs, then they would exit when their patronage dropped below some threshold level. In that case, while most switching consumers would be inframarginal and impose no negative externalities, the marginal consumer whose switching leads a store to shut down would impose potentially large negative externalities on its remaining would-be shoppers (in the spirit of [Waldfogel, 2008](#)). The net effect in these cases is theoretically ambiguous.

Unlike other studies of the retail sector, our analysis is not restricted to a small subset of retailers whose product composition is known and we do not have to guess at what stores actually sell. Using rich product-level sales data from the Census of Retail Trade micro files, we know what products stores sell, how much revenue each store generates from each product category, and how this varies both across stores in a single chain and at a single store over time. With these data, we document the simultaneous growth of general-merchandise chains and the expansion of the superstore format over a thirty-year period. This effect is particularly strong for large chains, consistent with their dramatic growth over these three decades.

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 (8), 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 Π has increasing differences in $(n, k; c_m, \tau, \delta)$, or equivalently, since Π 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 δ :

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

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 τ :¹⁹

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

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

¹⁹We 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:

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

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 Π with respect to k is

$$\frac{\partial \Pi}{\partial k} = \pi_s^* - \sum_{i=1}^n kx_s^i \frac{\partial}{\partial k} \left(\frac{w(kx_s^i)}{kx_s^i} \right) - \frac{1}{\delta} \frac{\partial d(n, k)}{\partial k} \quad (11)$$

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),

$$\frac{\partial^2 \Pi}{\partial k \partial n} = \frac{\partial \pi_s^*}{\partial n} + kx^* \cdot \left(\frac{w(kx^*)}{kx^*} - w'(kx^*) \right) - nk \frac{\partial x^*}{\partial n} w''(kx^*) - \frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k}. \quad (12)$$

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 (12) to be positive is that

$$\frac{\partial \pi_s^*}{\partial n} - \frac{1}{\delta} \frac{\partial^2 d(n, k)}{\partial n \partial k} > 0. \quad (13)$$

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 (13) holds at $n = k = 1$. This is guaranteed by the domain restriction $\delta > \bar{\delta}(R, c_m, \tau)$. \square

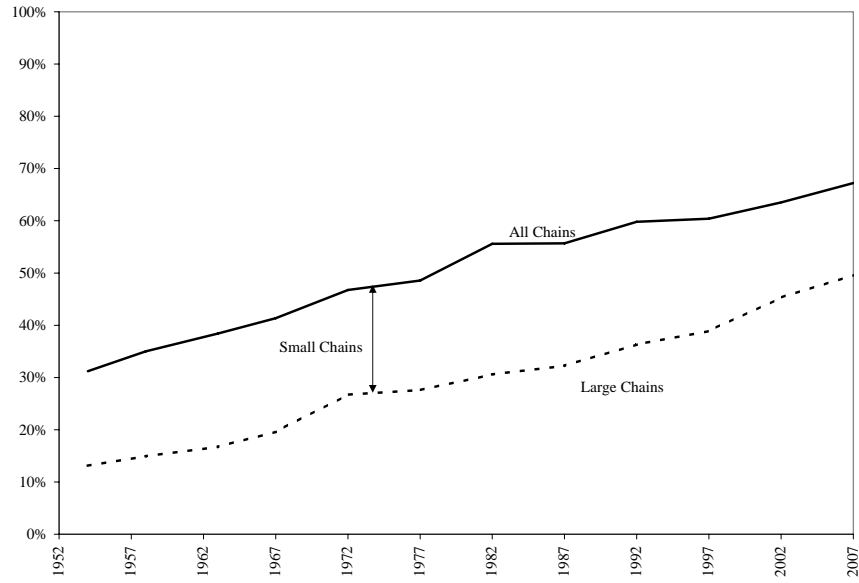
References

- Abernathy, F. H., J. T. Dunlop, J. H. Hammond, and D. Weil (1999) *A Stitch in Time: Lean Retailing and the Transformation of Manufacturing — Lessons from the Apparel and Textile Industries*. Oxford University Press, New York and Oxford.
- Acemoglu, D. (2002) “Directed Technical Change,” *Review of Economic Studies*, 69(4), 781–809.
- 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.
- (2011) “Raising the Barcode Scanner: Technology and Productivity in the Retail Sector,” unpublished paper, University of Missouri and U.S. Census Bureau.
- Basker, E., and M. Noel (2009) “The Evolving Food Chain: Competitive Effects of Wal-Mart’s Entry into the Supermarket Industry,” *Journal of Economics and Management Strategy*, 18(4), 977–1009.
- Basker, E., and P. H. Van (2008) “Wal-Mart as Catalyst to U.S.-China Trade,” unpublished paper, University of Missouri.
- (2010) “Imports ‘Я’ Us: Retail Chains as Platforms for Developing-Country Imports,” *American Economic Review Papers and Proceedings*, 100(2), 414–418.
- Beck, J., M. Grajek, and C. Wey (2011) “Estimating Level Effects in Diffusion of a New Technology: Barcode Scanning at the Checkout Counter,” *Applied Economics*, 43(14), 1737–1748.
- 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.
- Betancourt, R. R. (2005) *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, <http://www.msnbc.msn.com/id/4001039/4001039>.

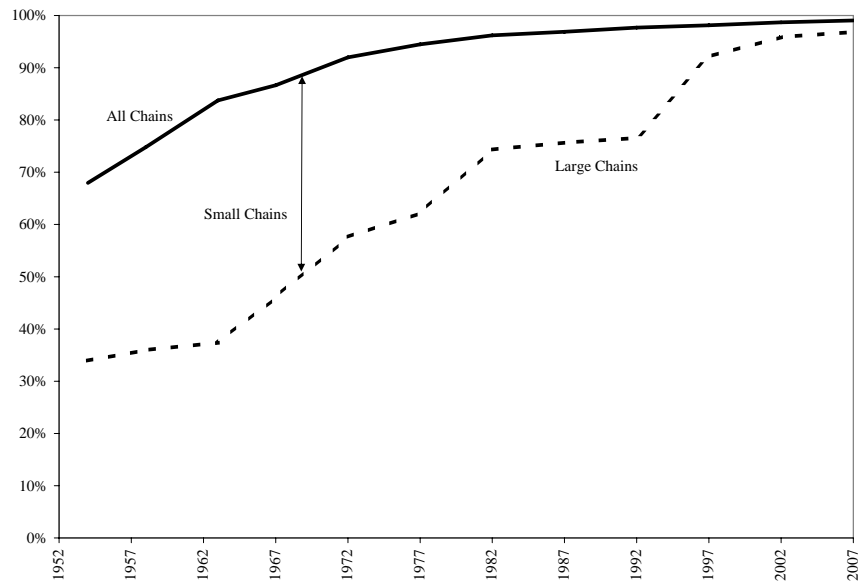
- Campbell, J. R., and T. N. Hubbard (2009) “The Economics of “Radiator Springs”: Industry Dynamics, Sunk Costs, and Spatial Demand Shifts,” Northwestern University Working Paper.
- Caprice, S., and V. von Schlippenbach (2008) “Competition Policy in a Concentrated and Globalized Retail Industry,” *Applied Economics Quarterly*, 54(3).
- Carden, A., and C. Courtemanche (2009) “Wal-Mart, Leisure, and Culture,” *Contemporary Economic Policy*, 27(4), 450–461.
- Chandler, A. D. (1969) *Strategy and Structure: Chapters in the History of the Industrial Enterprise*. MIT Press, Cambridge, MA.
- Christman, E. (2002) “Target and Wal-Mart Keep Rackjobbers Jumping,” *Billboard Magazine*, August 31.
- Das, N., E. M. Falaris, and J. G. Mulligan (2009) “Vintage Effects and the Diffusion of Time-Saving Technological Innovations: The Adoption of Optical Scanners by U.S. Supermarkets,” *B.E. Journal of Economic Analysis and Policy (Advances)*, 9(1), article 23.
- 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.
- Dukes, A. J., T. Geylani, and K. Srinivasan (2009) “Strategic Assortment Reduction by a Dominant Retailer,” *Marketing Science*, 28(2), 309–319.
- Dunlop, J. T., and J. W. Rivkin (1997) “Introduction,” in Stephen A. Brown, *Revolution at the Checkout Counter: The Explosion of the Bar Code*. Harvard University Press, Cambridge, MA.
- Ellickson, P. B. (2007) “Does Sutton Apply to Supermarkets?,” *RAND Journal of Economics*, 38(1), 43–59.
- Forman, C., A. Ghose, and A. Goldfarb (2009) “Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on where You Live,” *Management Science*, 55(1), 47–57.
- 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.
- Foster, L., J. Haltiwanger, and C. Syverson (2008) “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?,” *American Economic Review*, 98(1), 394–425.
- Gogoi, P. (2008) “Wal-Mart’s Crossover Strategy Still Works,” *Business Week*, March 7, http://www.businessweek.com/bwdaily/dnflash/content/mar2008/db2008036_169954.htm.

- Haltiwanger, J., R. Jarmin, and C. J. Krizan (2010) “Mom-and-Pop Meet Big Box: Complements or Substitutes?,” *Journal of Urban Economics*, 67(1), 116–134.
- Harmon, C. K., and R. Adams (1984) *Reading between the Lines: An Introduction to Bar Code Technology*. North American Technology, Inc., Peterborough, NH.
- Heher, A. (2009) “War of the Words,” *Chicago Daily Herald*, October 22.
- Holmes, T. (2001) “Bar Codes Lead to Frequent Deliveries and Superstores,” *RAND Journal of Economics*, 32(4), 708–725.
- (2011) “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1), 253–302.
- Hunter, S. D. (2003) “Learning, and the Market Value of the Firm,” MIT Sloan School of Management Working Paper 4418-03.
- Inderst, R., and C. Wey (2007) “Buyer Power and Supplier Incentives,” *European Economic Review*, 51(3), 641–667.
- Irwin, N. (2005) “Merger Approved, Sears-Kmart Can Focus on Reshaping Stores,” *Washington Post*, March 25.
- Jarmin, R. S., S. D. Klimek, and J. Miranda (2009) “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, pp. 237–262. University of Chicago Press.
- Jia, P. (2008) “What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Industry,” *Econometrica*, 76(6), 1263–1316.
- Jovanovic, B. (1982) “Selection and the Evolution of Industry,” *Econometrica*, 50(3), 649–670.
- Klemperer, P., and A. J. Padilla (1997) “Do Firms’ Product Lines Include Too Many Varieties?,” *RAND Journal of Economics*, 28(3), 472–488.
- Kosová, R., and F. Lafontaine (2010) “Survival and Growth in Retail and Service Industries: Evidence from Franchised Chains,” *Journal of Industrial Economics*, 58(3), 542–578.
- Lafontaine, F. (1992) “Agency Theory and Franchising: Some Empirical Results,” *RAND Journal of Economics*, 23(2), 263–283.
- Levin, S. G., S. L. Levin, and J. B. Meisel (1987) “A Dynamic Analysis of the Adoption of a New Technology: The Case of Optical Scanners,” *Review of Economics and Statistics*, 69(1), 12–17.
- Levine, R. (2004) “Would You Like an Extra Shot of Music With That Macchiato?,” *The New York Times*, November 3.

- Marvel, H. P., and J. Peck (2008) "Inventory Turnover and Product Variety," *Journal of Law and Economics*, 51(3), 461–478.
- Matsa, D. A. (forthcoming) "Competition and Product Quality in the Supermarket Industry," *Quarterly Journal of Economics*.
- 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>.
- Raff, D. M. G., and P. Temin (1999) "Sears, Roebuck in the Twentieth Century: Competition, Complementarities, and the Problem of Wasting Assets," in *Learning by Doing in Markets, Firms, and Countries*, ed. by N. R. Lamoreaux, D. M. G. Raff, and P. Temin, pp. 219–252. University of Chicago Press.
- Reuters (2002) "Analysis: Wal-Mart's Pain Should be Toys R Us Gain," *Rueters*, February 16.
- Rosenbloom, S. (2009) "Toys 'R' Us Trims Losses by Making a Hamster Hot," *The New York Times*, December 18.
- Singh, V. P., K. T. Hansen, and R. C. Blattberg (2006) "Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter," *Marketing Science*, 25(5), 457–476.
- Supermarket News (2009) "Report: Toys 'R' Us Explores Groceries," *Supermarket News*, April 3.
- Wal-Mart Stores, Inc. (various years) *Wal-Mart Annual Report*. Bentonville, AR.
- Waldfogel, J. (2008) "The Median Voter and the Median Consumer: Local Private Goods and Residential Sorting," *Journal of Urban Economics*, 68(2), 567–582.
- Yeap, C. (2011) "Capacity Use in Multi-Unit Firms: Evidence for Efficiency Gains or Strategic Competition in the U.S. Restaurant Industry?," *Applied Economics*, 43(22), 2953–2968.

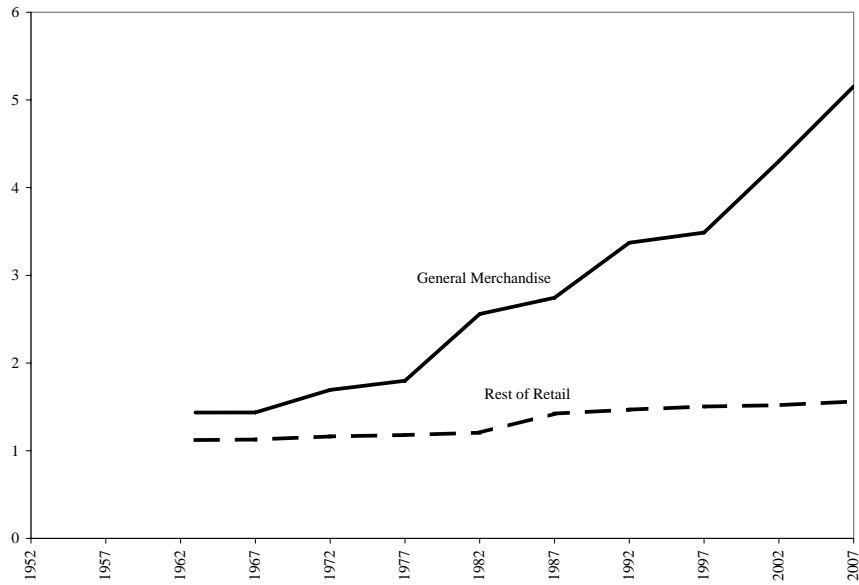


(a) All Retail

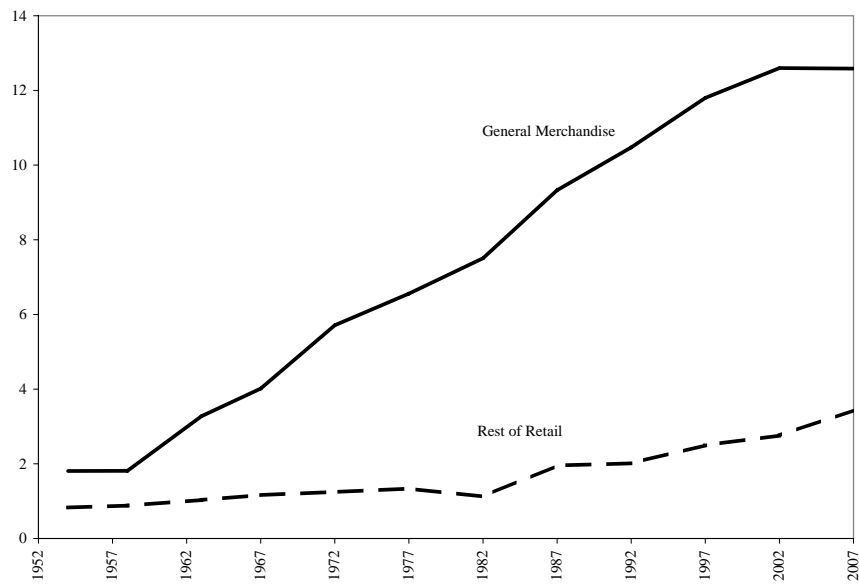


(b) General Merchandise

Figure 1. All Chains' and Large Chains' Share of Revenues
 Large chains are defined as chains with 100 or more stores
 Source: Published data from Census of Business and Census of Retail Trade



(a) Average Stores per Retailer



(b) Average Revenue per Store in Millions of 2007 Dollars

Figure 2. Chain and Store Growth over Time

Source: Published data from Census of Business and Census of Retail Trade

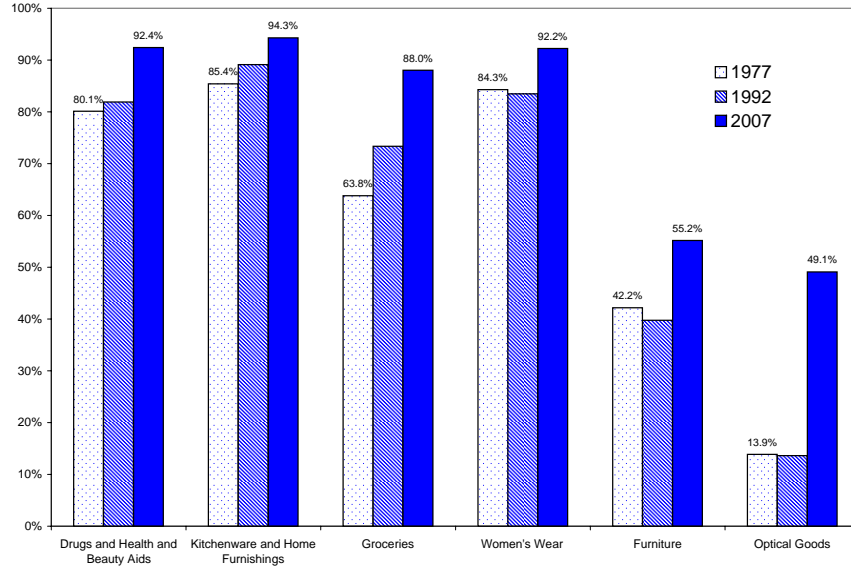


Figure 3. Share of General-Merchandise Establishments Selling Selected Products
Source: Published data from Census of Retail Trade

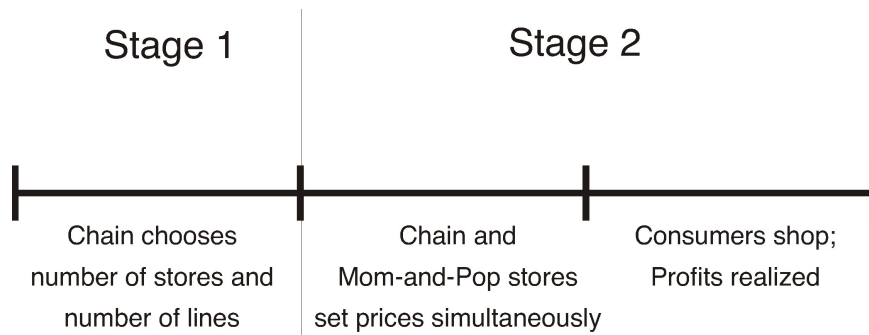
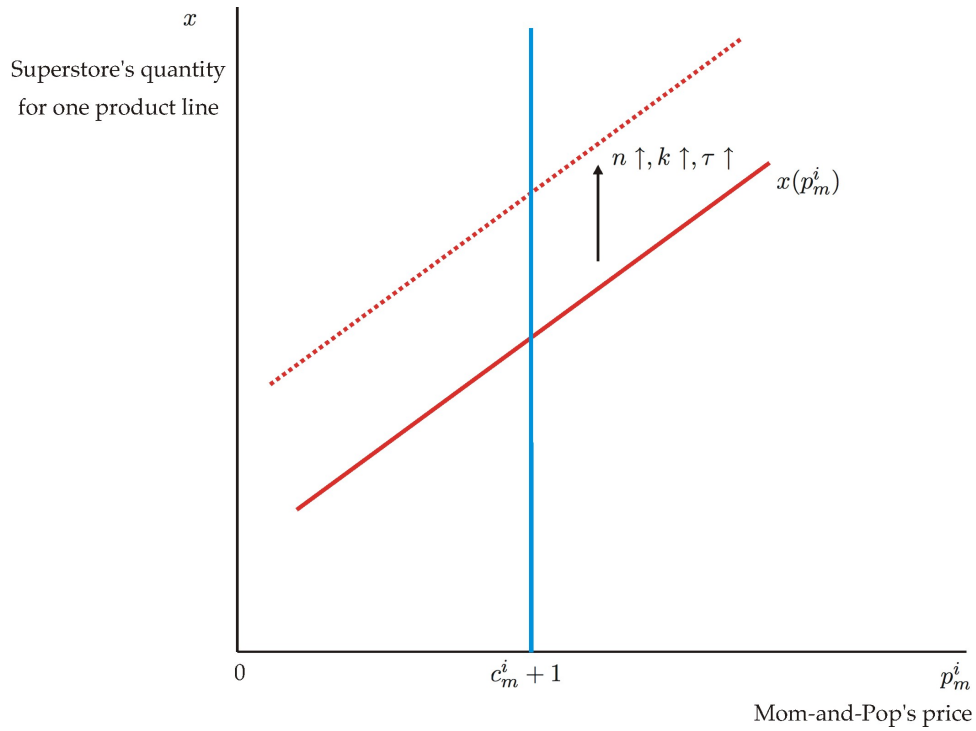
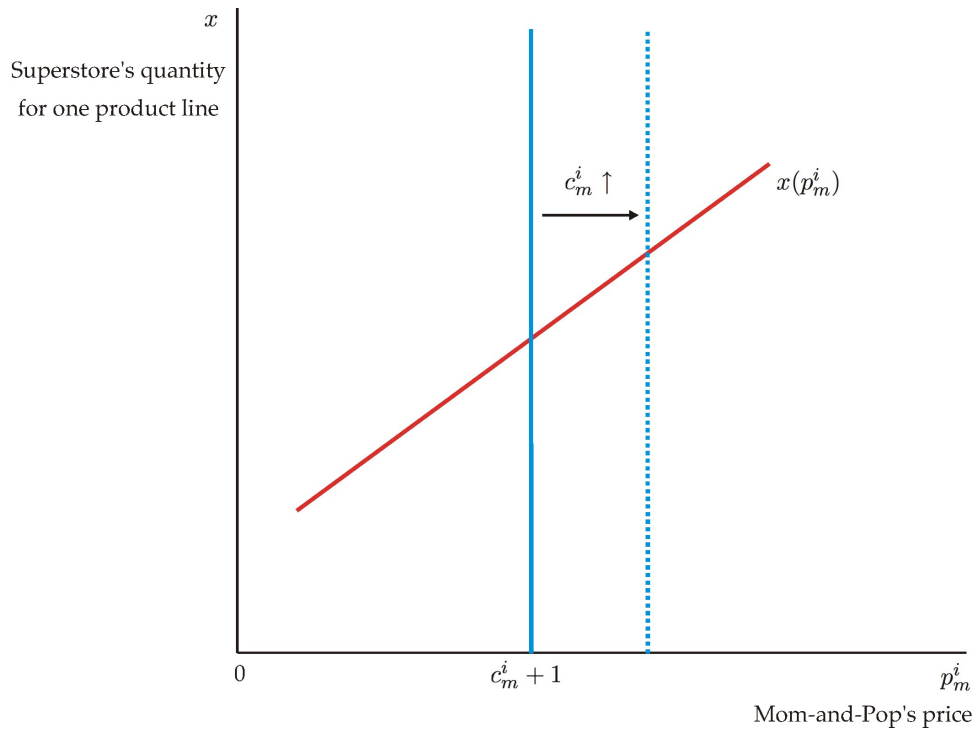


Figure 4. Timing of Model



(a) An increase in the number of product lines, n , number of stores, k , or transportation cost, τ .



(b) An increase in the mom-and-pop marginal cost, c_m^i .

Figure 5. Comparative Statics of the Second-Stage Nash Equilibrium

Table 1. General-Merchandise Subsector vs. Rest of Retail Sector, 2007

	General Merchandising ^a	Rest of Retail ^b	GM/All Retail
Retail Firms (000)	9	707	0.012
Stores (000)	46	1,082	0.041
Revenue (000,000,000\$)	577	3,341	0.147
Fraction of Stores in Chains	0.81	0.39	1.982
Fraction of Stores in Large Chains ^c	0.77	0.24	2.893

Source: Published data from Census of Retail Trade, 2007

^a NAICS 452

^b NAICS 44–45, excluding 452

^c Large chains have 100 or more stores

Table 2. The General-Merchandise Subsector over Time

Year	All Firms	Single Stores	Small Chains	Large Chains ^a	Total Revenues ^b
1977	16,098	14,897	1,166	35	320,776,187
1982	13,344	12,386	918	40	256,353,360
1987	12,917	12,093	786	38	330,628,856
1992	10,264	9,660	565	39	362,559,869
1997	10,373	9,933	398	42	426,884,830
2002	9,467	9,150	285	32	513,139,738
2007	8,925	8,616	281	28	577,098,195

Source: Published data from Census of Retail Trade, 1977–2007

^a Large chains have 100 or more stores (101+ in 1977)

^b Thousands of 2007 dollars

Table 3. Firm Micro Data Summary Statistics

A: All General-Merchandise Firms

Chain Size ^a	Observations	Firms ^b	Stores per Firm		Average Lines		Total Lines		Inverse Herfindahl		Share Dynamic ^c
					Broad	Detail	Broad	Detail	Broad	Detail	
Single Stores	14,442	11,379	1.0 (0.0)	12.0 (6.1)	16.8 (10.6)	12.0 (6.1)	16.8 (10.6)	5.7 (2.9)	8.0 (5.6)	0.38 (0.48)	
Small Chains	2,773	1,598	8.9 (14.6)	14.8 (6.3)	23.3 (12.9)	16.5 (7.1)	27.0 (15.5)	7.1 (3.2)	12.1 (7.7)	0.73 (0.45)	
Large Chains	235	88	627.7 (1,033.5)	21.0 (6.8)	36.2 (17.8)	27.6 (8.5)	53.7 (27.6)	10.0 (4.5)	20.1 (10.6)	0.93 (0.25)	
All	17,450	12,787	10.6 (139.9)	12.5 (6.3)	18.1 (11.5)	12.9 (6.8)	19.0 (13.1)	6.0 (3.1)	8.9 (6.4)	0.44 (0.50)	

B: All Dynamic General-Merchandise Firms^c

Chain Size ^a	Observations	Firms ^b	Stores per Firm		Average Lines		Total Lines		Inverse Herfindahl	
					Broad	Detail	Broad	Detail	Broad	Detail
Single Stores	5,430	2,367	1.0 (0.0)	13.5 (6.2)	19.8 (11.3)	13.5 (6.2)	19.8 (11.3)	6.2 (2.9)	9.0 (5.9)	
Small Chains	2,013	838	9.9 (15.6)	15.1 (6.3)	24.5 (13.3)	17.0 (7.1)	28.8 (16.0)	7.1 (3.2)	12.8 (8.0)	
Large Chains	219	72	659.0 (1,063.5)	21.1 (6.7)	37.1 (17.8)	27.8 (8.5)	54.7 (27.5)	10.0 (4.5)	20.6 (10.8)	
All	7,662	2,999	22.1 (210.1)	14.1 (6.4)	21.6 (12.5)	14.9 (7.1)	23.1 (15.0)	6.5 (3.2)	10.4 (7.1)	

Standard deviations in parentheses. Sample includes all general-merchandise firms reporting line sales, 1977–2007

^a Small chains have 2–99 stores; large chains have 100 or more stores.

^b Firms are counted more than once if they grow or shrink across size classes.

^c Dynamic firms show up in at least two Census years

Table 4. Panel Regressions Results: All Firms

	Average Lines		Total Lines		Inverse Herfindahl	
	Broad	Detail	Broad	Detail	Broad	Detail
ln(Stores)	0.0694*** (0.0215)	0.0722*** (0.0268)	0.1281*** (0.0199)	0.1508*** (0.0256)	0.0855*** (0.0192)	0.1167*** (0.0260)
Year FE	✓	✓	✓	✓	✓	✓
Coverage	✓	✓	✓	✓	✓	✓
Form × Year	✓	✓	✓	✓	✓	✓
Firm FE (count)	12,787	12,787	12,787	12,787	12,787	12,787
R^2	0.907	0.899	0.910	0.903	0.897	0.890
Observations	17,450	17,450	17,450	17,450	17,450	17,450

LHS variable: log number of lines sold by the firm.

Sample includes all general-merchandiser firms reporting line sales, 1977–2007.

Robust standard errors in parentheses (clustered by firm)

* significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5. Panel Regressions Results: Large Firms

	Average Lines		Total Lines		Inverse Herfindahl	
	Broad	Detail	Broad	Detail	Broad	Detail
ln(Stores)	0.1262*** (0.0436)	0.1326** (0.0650)	0.1504*** (0.0446)	0.1815*** (0.0613)	0.1186** (0.0571)	0.1447* (0.0843)
Year FE	✓	✓	✓	✓	✓	✓
Coverage	✓	✓	✓	✓	✓	✓
Form × Year	✓	✓	✓	✓	✓	✓
Firm FE (count)	88	88	88	88	88	88
R^2	0.860	0.830	0.817	0.803	0.906	0.802
Observations	235	235	235	235	235	235

LHS variable: log number of lines sold by the firm.

Sample includes general-merchandise firms with 100 or more stores reporting line sales, 1977–2007.

Robust standard errors in parentheses (clustered by firm)

* significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6. Panel Regressions Results: Mergers and Acquisitions

	Average Lines		Total Lines		Inverse Herfindahl	
	Broad	Detail	Broad	Detail	Broad	Detail
ln(Stores)	0.0870*** (0.0306)	0.0975*** (0.0354)	0.1456*** (0.0276)	0.1792*** (0.0338)	0.1202*** (0.0283)	0.1584*** (0.0361)
M&A	0.0640** (0.0318)	0.0912** (0.0410)	0.0767** (0.0310)	0.1092*** (0.0410)	0.0058 (0.0263)	0.0692* (0.0378)
Year FE	✓	✓	✓	✓	✓	✓
Coverage	✓	✓	✓	✓	✓	✓
Form × Year	✓	✓	✓	✓	✓	✓
Firm FE (count)	9,799	9,799	9,799	9,799	9,799	9,799
R ²	0.918	0.910	0.920	0.914	0.907	0.901
Observations	12,963	12,963	12,963	12,963	12,963	12,963

LHS variable: log number of lines sold by the firm.

Sample is all general merchandisers with lines data, 1982–2007.

Robust standard errors in parentheses (clustered by firm)

* significant at 10%, ** significant at 5%, *** significant at 1%.

Table B-1. Product Lines

Broad Line Description	Years	Number of Detailed Lines^a
Groceries	All	9
Meals, Snacks, and Nonalcoholic Beverages for Immediate Consumption	All	1
Meals, Snacks, and Beverages for Catered Events	2007	1
Packaged Liquor, Wine, and Beer	All	1
Tobacco Products and Accessories	All	1
Drugs and Health and Beauty Aids	All	6
Soaps, Detergents, and Household Cleaners	1987–2007	1
Paper and Related Products	1987–2007	1
Men's Apparel	All	11
Women's Apparel	All	13
Children's Apparel	All	4
Footwear	All	5
Curtains, Draperies, and Domestics	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	1987–2007	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	1987–2007	2
Souvenirs and Novelty Items, Including Seasonal Decorations	1997–2007	2
Books, Magazines, and Newspapers	All	2
Miscellaneous Merchandise, Not Elsewhere Classified	All	3
Non-merchandise Receipts	All	8

^aMaximum number. Depending on the year, there are up to 41 broad lines and up to 124 detailed lines.