Productivity Dynamics and the Role of "Big-Box" Entrants in Retailing*

Florin Maican†
and
Matilda Orth‡

October 29, 2009

Abstract

There are few attempts to use recent advances of structural estimation of sales generating functions on retail markets. We use a dynamic structural model to estimate total factor productivity in retail markets controlling for unobserved prices and local market characteristics. In particular, we assess whether entry of large ("big-box") stores drives exit and growth in the productivity distribution of incumbents. We use detailed data on all retail food stores in Sweden and control for endogeneity of large entrants by using political preferences in local markets. Our results suggest that it is important to consider nonlinearities in the productivity process, and to control for local market characteristics and selection when estimating retail productivity. Large entrants force low productive stores to exit and surviving stores to increase their productivity growth. The increase in growth declines with the productivity level of survivors, indicating that growth increases most among low productive survivors after large entry. Our findings suggest that large entrants play a crucial role for driving productivity growth.

Keywords: Retail markets; imperfect competition; industry dynamics; TFP; dynamic structural model.

JEL Classification: O3, C24, L11.

*We thank Daniel Ackerberg, Victor Aguirregabiria, Mats Bergman, Jan De Loecker, Pierre Dubois, Martin Dufwenberg, Lennart Hjalmarsson, Jordi Jaumandreu, Vincent Réquillart, Rune Stenbacka, Johan Stennek, Måns Söderbom, and seminar participants at Toulouse School of Economics and the University of Gothenburg for valuable comments and discussions. In addition, we thank participants at EEA 2008 (Milano), EARIE 2007 (Valencia), the Nordic Workshop in Industrial Organization 2007 (Stockholm), the Conference of the Research Network on Innovation and Competition Policy 2007 (Mannheim), and the Swedish Workshop on Competition Research 2007 (Stockholm) for helpful comments and suggestions. Special thanks to the Trade Union Institute for Economic Research (FIEF) and the Swedish Retail Institute (HUI) for providing the data. Financial support from the Swedish Competition Authority is gratefully acknowledged.

†Department of Economics, University of Gothenburg, Box 640, SE 405 30, Göteborg, Sweden, Phone +46-31-786 4866, Fax: +46-31-786 4154, *E-mail*: florin.maican@handels.gu.se

‡Department of Economics, University of Gothenburg, Box 640, SE 405 30, Göteborg, Sweden, Phone +46-31-786 5984, Fax: +46-31-786 4154, *E-mail*: matilda.orth@economics.gu.se

1 Introduction

Recent methods for structural estimation of production functions have almost only been applied to manufacturing industries.[1] There are few attempts to estimate multi-factor productivity in retail markets, where entry and exit have been found to play a more crucial role for labor productivity growth than in manufacturing (Foster et al. 2006). The major structural change in retail markets during the last few decades is in fact the entry of large ("big-box") stores, along with a drastic fall in the number of stores. The most striking example is the expansion of Wal-Mart, which has been found to greatly lower retail prices, and increase exit, of retail stores in the U.S., the "Wal-Mart effect".[2] For instance, the number of single-store retailers in the U.S. declined 55% from 1963 to 2002 (Basker 2007). Retail markets in Europe also follow the "big-box" trend, though on a smaller scale, with for example Carrefour, Metro, Schwartz, and Tesco. Although there is an emerging literature about retail markets, the impact of this structural change on productivity has not been given much attention.[3] Our goal is to combine recent extensions of the Olley and Pakes' (1996) framework to estimate total factor productivity (TFP) in retail markets, and to investigate the impact of increased competition from large entrants on exit and productivity growth. That is, to what extent do large entrants drive reallocation of inputs and outputs, i.e., exit of low productive stores and growth of surviving stores with different positions in the productivity distribution? Detailed data on all retail food stores in Sweden give us unique opportunities to analyze the questions at hand.

Productivity analysis in retailing is more complex than in many other industries because of the problem of measuring output (Griffith and Harmgart 2005, Reynolds et al. 2005). We use a dynamic structural model to estimate productivity, which has the advantage of allowing stores to have heterogenous responses to industry shocks (Ackerberg et al. 2007). In detail, our model is based on the following key features of retail markets: First, the most common characteristics of retail data are lumpy investments and a weak measure of intermediate inputs.[4] Because labor and capital are key inputs in retail markets, we follow Doraszelski and Jaumandreu (2009) and recover productivity from the optimal choice of labor. Second, because retail stores operate in local markets we control for local market characteristics, i.e. for large entrants and population density. We control for endogeneity of large entrants by using political preferences in local markets as instruments. Third, because large store types are more likely than smaller ones to survive larger shocks to productivity we control for selection, as do Olley and Pakes (1996). Fourth, recent literature emphasizes the importance of controlling for prices when estimating production functions in imperfect competitive markets (Foster et al. 2008, De Loecker 2009).

[1] Olley and Pakes (1996), Pavcnik (2002), Levinsohn and Petrin (2003), Ackerberg et al. (2006), Buettner (2004), De Loecker (2009), Doraszelski and Jaumandreu (2009).

[2] Basker (2005), Basker (2007), Basker and Noel (2009), Holmes (2008), and Jia (2008). Fishman (2006) and Hicks (2007) provide a general discussion on the Wal-Mart effect.

[3] Two European contributions are Bertrand and Kramarz (2002), who find that retail markets in France have lower labor growth and higher concentration as a consequence of regulation, and Sadun (2008), who finds that regulation in the UK reduces employment in independent stores.

[4] While Olley and Pakes (1996) assume strict monotonicity of the investment function to recover unobserved productivity, Levinsohn and Petrin (2003) use the intermediate input of materials.

Since store prices and quantities are rarely observed in retail data we control for unobserved prices by introducing a simple demand system as in Klette and Griliches (1996), and thus obtain mark-up estimates.[5] Compared to two-step estimators (Olley and Pakes 1996, Ackerberg et al. 2006), our one-step estimator has the advantages of increased efficiency and reduced computational burden. Identification comes from that we recover unobserved productivity from the labor demand function of known parametric form using a good measure of full-time adjusted wages. The assumption that labor is a static input abstracts from training, hiring and firing costs. We argue that this assumption is less restrictive in retail food than in many other industries because part time working is common, the share of skilled labor is low, and stores frequently adjust labor due to variation in customer flows.

The role of large entrants has a direct link to competition policy because the majority of OECD countries have entry regulations, though much more restrictive in Europe than in the U.S. The main rationale is that new entrants generate both positive and negative externalities, which require careful evaluation by local authorities. Advantages, such as productivity gains, lower prices, and wider product assortments, stand in contrast to drawbacks, in terms of fewer stores, and environmental issues. Because we anticipate large entrants to have an extensive impact on market structure, they are carefully evaluated in the planning process. The consequences of regulation (e.g. supermarket dominance) are frequently debated among policy makers in Europe (European Parliament 2008). Our primary objective is not to quantify the magnitude of inter-firm reallocations over time, i.e., how (large) entrants, exits, and incumbents contribute to aggregate productivity growth.[6] Instead we provide reduced form evidence for how large entrants influence exit and the productivity growth of incumbents in local markets.

We focus on food retailing because it accounts for a large (15%) share of consumers' budgets (Statistics Sweden 2005) and thus constitutes a large share of retailing. Besides, many other service sectors follow similar trends as retail food. The Swedish market is appropriate to analyze because it follows two crucial trends common among nearly all OECD countries: There has been a structural change towards larger but fewer stores; in fact, the total number of stores in Sweden declined from 36,000 in the 1950s to below 6,000 in 2003 (Swedish National Board of Housing, Building, and Planning 2005). And there is an entry regulation that gives municipalities power to decide over the land use and, consequently, whether or not a store is allowed to enter the market.

Our study connects to the literature of dynamic models with heterogenous firms (Jovanovic 1982, Hopenhayn 1992, and Ericson and Pakes 1995). In particular, we build on the growing literature on productivity heterogeneity within industries that use dynamic structural models (Olley and Pakes 1996, Pavcnik 2002, Levinsohn and Petrin 2003, Ackerberg et al. 2006, Acker-

[5] Other studies that introduce prices in the production function are Melitz (2000), Levinsohn and Melitz (2002), Katayama et al. (2003), and Doraszelski and Jaumandreu (2009). In contrast to Doraszelski and Jaumandreu, who account for observed prices, we account for unobserved store prices.

[6] We estimate the contribution of *all* entrants to aggregate productivity growth using various TFP decompositions (Griliches and Regev 1995, Foster et al. 2001, Melitz and Polanec 2009) but, due to data constraints, we cannot quantify the exact contribution of large entrants.

berg et al. 2007). We instead back out productivity from labor demand , as do Doraszelski and Jaumandreu (2009). Recent studies on productivity dynamics show two important facts: large and persistent productivity differences among firms, and substantial reallocation across firms in the same industry.[7] They found that the key mechanism to foster growth is reallocation from less to more productive firms, either through less productive firms exiting and more productive firms entering or through increased productivity among incumbents, or both. Thus, increased competition forces low productive firms to exit, increasing the market shares of more productive firms. The productivity distribution is thus truncated from below, increasing the mean, and decreasing dispersion (Melitz 2003, Asplund and Nocke 2006). Using a local market approach, Syverson (2004) emphasizes that demand density result in similar improvements in productivity distribution. In retail, entry and exit have been found to contribute to almost all labor productivity growth in the U.S., where chains replace low productive firms with high productive entrants (Foster et al. 2006). In Sweden, large food stores have been found to offer lower prices than others (Asplund and Friberg 2002). However, how large entrants influence local market competition and changes in the productivity distribution of incumbents has not been analyzed in detail.[8]

The empirical results show that it is important to allow for nonlinearities in the productivity process and to control for prices, local markets characteristics, and selection when estimating retail productivity. Large entrants force low productive stores to exit, and surviving stores to increase their productivity growth. Growth increases most among incumbents in the bottom part of the productivity distribution, and then declines with the productivity level of incumbents. Large entrants thus spur reallocation of resources towards more productive stores. From a policy perspective, we claim that a more liberal design and application of entry regulations would support productivity growth in retail markets.

The next section describes the retail food market and the data. Section 3 presents the modeling approach for estimating productivity, and those results. Section 4 reports the link between large entrants and exit and productivity growth. Section 5 summarizes and draws conclusions.

2 The retail food market and data

Retail food markets in the OECD countries are fairly similar, consisting of firms operating uniformly designed store types. In Sweden, the food market is dominated by four firms that together had 92% of the market shares in 2002: ICA(44%), Coop(22%), Axfood(23%), and

[7]Caves (1998) and Bartelsman and Doms (2000) provide surveys, mainly on manufacturing.

[8]The paper also relates to the vast literature on how competition affects productivity, emphasizing both positive and negative effects theoretically, but often positive effects empirically. Recent theoretical contributions are Nickell (1996), Schmidt (1997), Boone (2000), Melitz (2003), and Raith (2003); whereas recent empirical contributions include Porter (1990), MacDonald (1994), Nickell (1996), Blundell et al. (1999), Aghion et al. (2003), Sivadasan (2004), and Aghion et al. (2009).

Bergendahls(3%). Various independent owners make up the remaining 8% market share.[9] ICA consists mostly of independently owned stores with centralized decision making. Coop, on the other hand, consists of centralized cooperatives with decisions made at national or local level. Axfood and Bergendahls each have a mix of franchises and centrally owned stores, the latter mainly in the south and southwest of Sweden.[10]

A majority of OECD countries have entry regulations that give power to local authorities. The regulations differ substantially across countries, however (Hoj et al. 1995, Boylaud and Nicoletti 2001, Griffith and Harmgart 2005, Pilat 2005). While some countries strictly regulate large entrants, more flexible zoning laws exist, for instance, in the U.S. (Pilat 1997). The Swedish Plan and Building Act (PBA) gives power to the 290 municipalities to decide over applications for new entrants. In case of inter-municipality questions of entry, they are handled by the 21 county administrative boards. PBA is claimed to be one of the major barrier to entry, resulting in diverse outcomes, e.g., in price levels, across municipalities (Swedish Competition Authority 2001:4). Several reports stress the need to better analyze how regulation affects market outcomes (Pilat 1997, Swedish Competition Authority 2001:4, 2004:2). Large entrants are often newly built stores in external locations, making regulation highly important.[11] Appendix A describes PBA in greater detail.

■ **Data.** In order to cover various store productivity measures and define large entrants, we use two micro-data sets. The first data set, collected by Delfi Marknadsparter AB (DELFI), defines a unit of observation as a store based on its geographical location, i.e., its physical address (sources are described in Appendix A). This data, covering all retail food stores in the Swedish market during 1995-2002, include store type, chain, revenue class, and sales space (in square meters). The store type classification (12 different) depends on size, location, product assortment etc. An advantage with DELFI is that it contains all stores and their physical locations; shortcomings are a lack of input/output measures and the fact that revenue information is collected by surveys and reported in classes. Therefore, we use DELFI only to define large entrants.

The most disaggregated level for which more accurate input and output measures exist is organization number (Statistics Sweden, SCB). SCB provides data at this level based on tax reporting. But due to anonymous codes, the two data sets cannot be linked. Financial Statistics (FS) provides input and output measures, and Regional Labor Statistics (RAMS) comprises data on wages for all organization numbers from 1996 to 2002 belonging to SNI code 52.1, "Retail sales in non-specialized stores", which covers the four dominant firms (ICA, Coop, Axfood, and

[9] International firms with hard discount formats entered the Swedish market after the study period: Netto in 2002, and Lidl in 2003.

[10] In 1997, Axel Johnson and the D-group merged, initiating more centralized decision making and more uniformly designed store concepts.

[11] Possibly, firms can adopt similar strategies as their competitors and buy already established stores. As a result, more productive stores can enter without involvement of PBA and, consequently, the regulation will not work as an entry barrier that potentially affects productivity. Of course, we cannot fully rule out the opportunity that firms buy already established stores.

Bergendahls).[12] This FS-RAMS data, at the organization number level, consist of "multi-store" units, which may be one store or more with the same organization number (e.g., due to having the same owner).[13] Over 80% of the stores in DELFI each have their own organization number, so that less than 20% of the observations in FS-RAMS consist of two or more stores (discussed in detail below). If a firm consists of two stores, we observe total, not average, inputs and outputs. Note that all stores are reported in both data sets. Appendix A gives more information about the FS-RAMS data. Finally, we connect demographic information (population, population density, average income, and political preferences) from SCB to FS-RAMS and DELFI.

■ **Local markets.** Food products fulfill daily needs, are often of relatively short durability, and stores are thus located close to consumers. The travel distance when buying food is relatively short (except if prices are sufficiently low), and nearness to home and work are thus key aspects for consumers choosing where to shop, though distance likely increases with store size.[14] The size of the local market for each store depends on its type. Large stores attract consumers from a wider area than do small stores, but the size of the local market also depends on the distance between stores. We assume that retail markets are isolated geographic units, with stores in one market competitively interacting only with other stores in the same local market. A complete definition of local markets requires information about the exact distance between stores. Without this information we must rely on already existing measures. The 21 counties in Sweden are clearly too large to be considered local markets for our purposes, while the 1,534 postal areas are probably too small, especially for large stores (on which we focus). Two intermediate choices are the 88 local labor markets or the 290 municipalities. Local labor markets take into account commuting patterns, which are important for the absolutely largest types such as hypermarkets and department stores, while municipalities seem more suitable for large supermarkets. As noted, municipalities are also the location of local government decisions regarding new entrants. We therefore use municipalities as local markets.

■ **Large entrants and endogeneity.** DELFI relies on geographical location (address) and classifies store types, making it appropriate for defining large entrants. Because of a limited number of large stores, we need to analyze several of the largest store types together. We define the five largest types (hypermarkets, department stores, large supermarkets, large grocery stores, and other[15]) as "large" and four other types (small supermarkets, small grocery stores, convenience stores, and mini markets) as "small".[16] Gas station stores, seasonal stores, and

[12]SNI (Swedish National Industry) classification codes build on the EU standard NACE.

[13]FS-RAMS does not rely on addresses like DELFI, so we could not do a more detailed investigation of TFP and geographical distance (location).

[14]The importance of these factors is confirmed by discussions with representatives from ICA, COOP, and Bergendahls. According to surveys made by the Swedish Institute for Transport and Communication Analysis, the average travel distance for trips with the *main* purpose of buying retail food products is 9.83 kilometers (1995-2002).

[15]Stores classified as other stores are large and externally located.

[16]Alternatively, we define observations in FS-RAMS with sales above the 5th percentile of large stores' sales in DELFI as large; otherwise as small. Even though the available data do not allow a perfect match, the number of large entrants in FS-RAMS (so defined) follows a trend over time similar to that of the large entrants in DELFI. The empirical results (available from the authors upon request) are consistent with those reported here.

stores under construction are excluded due to these types not belonging in the SNI-code 52.1 in FS-RAMS. From the point of view of the Swedish market, we believe that these types are representative of being large.

A key problem when analyzing the link between large entrants and productivity growth is the endogeneity of large entry. We hence need to bring exogenous variation in large entry using instruments. No major policy reforms changing the conditions for large entrants have taken place in Sweden during the study period (see Appendix A for details about PBA).[17] Local authorities in Sweden decide however about entry of big-box stores. Following Bertrand and Kramarz (2002) and Sadun (2008) we use political preferences in municipalities as instruments for large entrants.[18] We use variation in political preferences across local markets throughout the election periods 1994-1998 and 1999-2002 to add exogenous variation in the number of large entrants. We expect non-socialist local governments to have a more liberal view of large entrants, though the number of applications and rejections to each municipality is unfortunately not available in Sweden. Our instruments are valid if political preferences capture decision-making about large entrants and are uncorrelated with unobserved shocks.

■ **Descriptive statistics.** Table 1 presents descriptive statistics of the Swedish retail food industry from the two data sets DELFI and FS-RAMS for 1996-2002. As noted, over 80% of the observation units in FS-RAMS are identical to the stores in DELFI. The rest (20% in the beginning and 14% in the end) are multi-store units in FS-RAMS. The number of stores in DELFI decreases over the period from 4,664 to 3,585, i.e., a reduction of over 23%, indicating that many stores closed. In FS-RAMS, the number of observations decreases by about 17% (from 3,714 to 3,067).[19] The share of large stores in DELFI increases from 19% to nearly 26%. While total sales space is virtually constant, mean sales space increases by 33%. Thus there has been a major structural change towards larger but fewer stores in the Swedish retail food market. Total wages (in FS-RAMS) increase over 22% (in real terms), while the number of employees increases only 9%.[20] Total sales increase about 26% (in FS-RAMS). Total sales in DELFI are lower and increase only 10% due to survey collection and interval reporting.

Table 2 shows median characteristics of local markets (municipalities) with and without large entrants during 1997-2002. The median number of stores varies between 22 and 54 in large entry markets, compared to 13-15 in non-entry markets. The number of markets with at least one large entrant varies between 6 and 23. Among these, up to 3 large entrants establish in the same market in the same year. As we expect, median entry and exit are higher in large entry than in non-entry markets, with median population, population density, and income also higher there. Large entry markets also have a lower concentration; the median four store concentration ratio is about 0.5 in these markets while it is over 0.7 in markets without large entrants.

[17] Studies based on UK data have used major policy reforms to handle endogeneity of entry (Sadun 2008, Aghion et al. 2009).

[18] Data on the number of formal applications for entry, and rejections, is unfortunately not available in Sweden.

[19] This indicates that entry and exit based on changes in organization numbers in FS-RAMS in some cases differ from entry and exit based on addresses in DELFI due to, e.g., re-organizations.

[20] The aggregate growth of real wages in Sweden was 24% during the period.

3 Productivity estimation

Our model of competition among retail stores is based on Ericson and Pakes' (1995) dynamic oligopoly framework. A store is described by a vector of state variables consisting of productivity $\omega \in \Omega$, capital stock $k \in \mathbb{R}_+$, and local market characteristics $z \in Z$. Incumbent stores maximize the discounted expected value of future net cash flows. Stores compete in the product market and collect their payoffs. At the beginning of each time period, incumbents decide whether to exit or continue to operate in the local market.[21] If the store exits, scrap value ϕ is received. If the store continues, it chooses optimal levels of labor l and investment $i \geq 0$. We assume capital is a dynamic input that accumulates according to $k_{t+1} = (1-\delta)k_t + i_t$, where δ is the discount rate. Labor, on the other hand, is a non-dynamic input chosen based on current productivity. Changes in investment do not guarantee a more favorable state tomorrow, but do guarantee more favorable distributions over future states. As in Olley and Pakes (1996)(hereafter OP), the transition probabilities of productivity follow a first order Markov process with $P(d\omega|\omega)$. We denote $V(\omega_{jt}, k_{jt}, z_{mt})$ to be the expected discounted value of all future net cash flows for store j in market m at period t. $V(\omega_{jt}, k_{jt}, z_{mt})$ is defined by the solution to the following Bellman equation with the discount factor $\beta < 1$

$$\begin{aligned} V(\omega_{jt}, k_{jt}, z_{mt}) = \quad & max\left\{\phi, \sup_{i_{jt}}[\pi(\omega_{jt}, k_{jt}, z_{mt}) - c_i(i_{jt}, k_{jt}) - c_l(l_{jt}) + \right. \\ & \left. \beta E[V(\omega_{jt+1}, k_{jt+1}, z_{mt+1})|\omega_{jt}, i_{jt}]\right\} \end{aligned} \tag{1}$$

where $\pi(\omega_{jt}, k_{jt}, z_{mt})$ is the profit function, increasing in both ω_{jt} and k_{jt}; $c_i(i_{jt}, k_{jt})$ is investment cost in new capital, where $c_i(\cdot)$ is increasing in investment choice i_{jt} and decreasing in capital stock k_{jt}; and $c_l(l_{jt})$ is the cost of labor, where $c_l(\cdot)$ is increasing in labor l_{jt}. Incumbent stores are assumed to know their scrap value ϕ prior to making exit and investment decisions. The solution of the store's optimization problem (1) gives optimal policy functions for labor $l_{jt} = \tilde{l}_{jt}(\omega_{jt}, k_{jt}, z_{mt})$, investment $i_{jt} = \tilde{i}_{jt}(\omega_{jt}, k_{jt}, z_{mt})$, and exit decision $\chi_{jt+1} = \tilde{\chi}_{jt}(\omega_{jt}, k_{jt}, z_{mt})$. The exit rule χ_{jt+1} depends on the threshold productivity $\underline{\omega}_{mt}(k_{jt}, z_{mt})$, where z_{mt} is a vector of local market characteristics such as the number of large entrants e^L_{mt}, and population density p^{dens}_{mt}.

■ **Value added generating function.** We assume Cobb-Douglas technology where stores sell a homogeneous product, and that the factors underlying profitability differences among stores are neutral efficiency differences.[22] We follow the common notation of capital letters for levels

[21] The decision to exit or continue is made at the store level, though firms that operate several stores can influence the decision of each store through possible chain effects. However, the firm takes the decision based on store performance.

[22] We can easily apply another specification; for example, translog with neutral efficiency across stores would do equally well.

and small letters for logs. The production function can be specified as

$$q_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + u^p_{jt} \tag{2}$$

where q_{jt} is the log of quantity sold by store j at time t; l_{jt} is the log of labor input; and k_{jt} is the log of capital input. The unobserved ω_{jt} is productivity, and u^p_{jt} is either a measurement error (which can be serially correlated) or a shock to productivity that is not predictable during the period in which labor can be adjusted. Since physical output is complex to measure in retail markets and therefore not observed, we use deflated value added as a proxy for output.

■ **Imperfect competition.** Equation (2) assumes that prices are constant across stores.[23] Foster et al. (2008) analyze the relation between physical output, revenues, and firm-level prices in the context of market selection. They find that productivity based upon physical quantities is negatively correlated with establishment-level prices, but productivity based upon revenues is positively correlated with those prices. The retail food market is characterized by imperfect competition, and product differentiation is a key factor. When a store has some market power, its price influences its productivity. If a store cuts its price, then more inputs are needed to satisfy increasing demand. This negative correlation between inputs and prices leads to underestimation of the labor and capital parameters in the production function (Klette and Griliches 1996, Melitz 2000, De Loecker 2009).[24] Following this literature, we consider a standard horizontal product differentiation demand system

$$p_{jt} = p_{mt} + \frac{1}{\eta} q_{jt} - \frac{1}{\eta} q_{mt} - \frac{1}{\eta} \lambda_{jt} - \frac{1}{\eta} u^d_{jt} \tag{3}$$

where p_{jt} is output price, p_{mt} and q_{mt} are output price and quantity in local market m, λ_{jt} is demand shifters (observed and unobserved), and u^d_{jt} is a simple i.i.d. shock to demand. The parameter η (< -1 and finite) captures the elasticity of substitution among stores.[25] Due to data constraints the demand system is quite restrictive, implying a single elasticity of substitution for all stores, so that there are no differences in cross price elasticities, i.e., we have a constant markup over marginal cost $(\frac{\eta}{1+\eta})$, and the Learner index is $(\frac{1}{|\eta|})$. We can however allow the elasticity of substitution to differ across local market groups such as counties (21 in total). The Learner index for county g is then $\frac{1}{|\eta_g|}$.

We decompose demand shifters λ_{jt} into observed local market characteristics z_{mt}, i.e., number of large entrants, population density, and unobserved demand shocks υ_{jt} as

$$\lambda_{jt} = z'_{mt}\beta_z + \upsilon_{jt}$$

[23] Under perfect competition, productivity of the price-taking stores is not influenced by store level prices.

[24] If the products are perfect substitutes, then deflated sales are a perfect proxy for unobserved quality adjusted output.

[25] The vertical dimension is to some extent also captured since deflated output measures both quantity and quality, which is correlated with store type (size).

where υ_{jt} are either correlated unexpected shocks to demand or i.i.d. The unobserved demand shocks υ_{jt} are unobserved by the econometrician but known to or predictable by the stores when they make their input, price or exit decisions.

Since we have unobserved store prices and quantities, we use the deflated value added y_{jt}, defined as $q_{jt} + p_{jt} - p_{mt}$, as output in the estimation of the sales (value added) generating function. However, if p_{mt} is unobserved, the consumer price index for food products p_{It} can be used as a proxy. Controlling for unobserved store price p_{jt} in the value added generating function in (2), we then have

$$y_{jt} \equiv \left(1+\frac{1}{\eta}\right)[\beta_0 + \beta_l l_{jt} + \beta_k k_{jt}] - \frac{1}{\eta} q_{mt} - \frac{1}{\eta} z_{jt}^{'} \beta_z + \left(1+\frac{1}{\eta}\right)\omega_{jt} - \frac{1}{\eta}\upsilon_{jt} - \frac{1}{\eta} u_{jt}^d + \left(1+\frac{1}{\eta}\right) u_{jt}^p \quad (4)$$

Assuming that store productivity follows an exogenous first order Markov process, actual productivity can be written as the sum of expected productivity given the store information set $\mathcal{F}_{t-1}$, $E[\omega_{jt}|\mathcal{F}_{t-1}]$, and the i.i.d. productivity shock ξ_{jt}

$$\omega_{jt} = E[\omega_{jt}|\mathcal{F}_{t-1}] + \xi_{jt}. \quad (5)$$

The conditional expectation function $E[\omega_{jt}|\mathcal{F}_{t-1}]$ is unobserved by the econometrician (though known to the store). The shock ξ_{jt} may be thought of as the realization of uncertainties that are naturally linked to productivity. Therefore, the value added generating function becomes

$$\begin{aligned} y_{jt} = \; & \left(1+\tfrac{1}{\eta}\right)[\beta_0 + \beta_l l_{jt} + \beta_k k_{jt}] - \tfrac{1}{\eta} q_{mt} - \tfrac{1}{\eta} z_{jt}^{'} \beta_z + \left(1+\tfrac{1}{\eta}\right) E[\omega_{jt-1}|\mathcal{F}_{t-1}] + \\ & \left(1+\tfrac{1}{\eta}\right)\xi_{jt} - \tfrac{1}{\eta}\upsilon_{jt} - \tfrac{1}{\eta} u_{jt}^d + \left(1+\tfrac{1}{\eta}\right) u_{jt}^p \end{aligned} \quad (6)$$

Due to unobserved prices at the store level, we face a trade-off between a flexible approximation of ω_{jt} and separation of demand shocks from productivity.[26] The estimation strategy chosen depends on whether demand shocks υ_{jt} are thought to be correlated over time and on whether we use a linear or nonlinear approximation of the conditional expectation $E[\cdot]$ (Ackerberg et al. 2007). We first present Case (1) when υ_{jt} is correlated over time, which includes ω_{jt} and υ_{jt} following either a general Markov process or an AR(1). The Markov processes can be either dependent or independent. Under AR(1), ω_{jt} and υ_{jt} can follow either the same or different processes and no further assumptions are needed to estimate the parameters. Then we present Case (2) when υ_{jt} is i.i.d.

Case (1): *υ_{jt} are correlated over time*

First, if ω_{jt} and υ_{jt} follow dependent Markov processes then υ_{jt-1} will enter as a separate variable in the conditional expectation $E[\omega_{jt}|\omega_{jt-1}, \upsilon_{jt-1}]$. To solve the identification problem in (6) we need an estimate of υ_{jt-1}. The Berry et al. (1995) (BLP) literature produces estimates

[26] The alternative of not controlling for prices at all requires even stronger assumptions.

of a set of "unobserved product characteristics" that might be used as υ_{jt} (Ackerberg et al. 2007 discuss this in detail), which we might interpret as unobserved store quality. But in our case, due to the data constraint, it is impossible to back out υ_{jt} using this method.

Second, if ω_{jt} and υ_{jt} follow independent Markov processes then expected productivity at time t conditional on information set $\mathcal{F}_{t-1}$ does not depend on υ_{jt-1}. But in this case υ_{jt} is an important determinant of optimal labor or investment, and thus affects actual productivity. Since we have two unobservables (ω_{jt} and υ_{jt}) and no other control variable for υ_{jt}, identification in (6) requires one of the following assumptions:

(a) $\tilde{\omega}_{jt} \equiv (1+\frac{1}{\eta})(\omega_{jt} - \frac{1}{\eta}\upsilon_{jt})$, i.e., quality adjusted productivity, follows a first order nonlinear Markov process: $\tilde{\omega}_{jt} = E[\tilde{\omega}_{jt}|\mathcal{F}_{t-1}] + \xi_{jt} = \tilde{h}(\tilde{\omega}_{jt-1}) + \xi_{jt}$, where $\tilde{h}$ is an approximation of the conditional expectation (Melitz 2000, Levinsohn and Melitz 2002). In other words, a positive shock in either productivity or demand makes the store sell more but the exact source of the shock does not matter.

(b) ω_{jt} and υ_{jt} follow different AR(1) processes.[27] We assume that $\omega_{jt} = \rho_1\omega_{jt-1} + \xi_{jt}$ and $\upsilon_{jt} = \rho_2\upsilon_{jt-1} + \mu_{jt}$. One way to eliminate the unobserved demand shock from the value added generating function (6) is to take the first difference $\tilde{y}_{jt} = y_{jt} - \rho_1 y_{jt-1}$. If $\rho_1 = \rho_2$, this is sufficient for identification. If $\rho_1 \neq \rho_2$, the unobserved demand shock υ_{jt} is completely removed if we apply the difference $\tilde{y}_{jt} - \rho_2\tilde{y}_{jt-1}$ in (6). Note that $\tilde{y}_{jt} - \rho_2\tilde{y}_{jt-1}$ is stationary if $\rho_1 > \rho_2$, i.e., if productivity is more persistent than the demand shock (the roots of $\tilde{y}_{jt} - \rho_2\tilde{y}_{jt-1}$ are $\rho_2 - \rho_1$ and $-\rho_2$).

The advantage of (a) is that it allows for nonlinearity in the productivity process and the possibility of controlling for selection (see Case (2)). The drawbacks of (a) are that we observe quality-adjusted productivity and that we need more assumptions to back out productivity. The advantage of (b) is that we can sort out persistent demand shocks from productivity and that no more assumptions are needed for identification. A drawback of allowing for two AR(1) processes in (b) is that it is more data demanding, because we need two lags, and thus dropping two years of data, to make sure that we have removed the persistent unobserved demand shocks. Since a store needs to be present in the data for at least three years, this severely restricts the dynamics.

Case (2) υ_{jt} *are i.i.d.*

In this case, demand shocks are not correlated with inputs or with exit decisions. Therefore υ_{jt} collapses into the i.i.d. demand shocks from the price equation u^d_{jt}. Below we describe the estimation strategy when productivity follows a general Markov process.

■ **Inverse labor demand function.** A central feature of retail data is lumpy investment and a weak measure of intermediate inputs. Labor and capital are thus the two key inputs. Following

[27] See the dynamic panel model of Blundell and Bond (2000).

Doraszelski and Jaumandreu (2009), we recover productivity from the optimal choice of labor. We use a good measure of firm specific wages adjusted for full time hours.[28] The idea relies on Levinsohn and Petrin (2003) who recover unobserved productivity from the demand for static intermediate input of materials. We assume that labor is a static and variable input chosen based on current productivity. The functional form of the value added generating function provides a parametric form of the labor demand function, unlike Levinsohn and Petrin (2003) and Ackerberg et al. (2006) that are non-parametric in materials. The advantage is that we can include many stores with zero investment while not making any assumptions about the stores' dynamic programming problem. In abstract of store level wages it may however be hard to estimate the coefficients of static inputs in the Cobb-Douglas case (Bond and Söderbom 2005).

Our assumption that labor is a static and variable input abstracts from costs of training, hiring and firing employees, though for several reasons this is less restrictive in retail than in many other industries. Part time workers are common. As much as 40% of the employees in retail food work part time, compared to 20% for the Swedish economy as a whole (Statistics Sweden). The share of skilled labor is lower. Only 15% of the retail employees had a university education in 2002, compared to 32% for the total Swedish labor force (Statistics Sweden). Stores have long opening hours and adjust their labor due to variations in customer flows over the day, week, month and year. Finally, the training process might be shorter than in many other industries.

Our assumption that each store chooses labor based on its productivity implies that labor l_{jt} is correlated with the random productivity shock ξ_{jt}. In year t, stores chose current labor l_{jt} based on current productivity ω_{jt}, which gives labor demand as

$$l_{jt} = \frac{1}{1-\beta_l}[\beta_0 + ln(\beta_l) + \alpha + \beta_k k_j + \omega_{jt} - (s_{jt} - p_{jt})] \tag{7}$$

where $\alpha = lnE[e^{\xi_{jt}}]$ and s_{jt} is log of total wages paid by store j in period t. Under the functional form assumption on the value added generating function, we have a known functional form for the labor demand and inverse labor demand functions. Solving for ω_{jt} in Equation (7) yields the inverse labor demand function from which we can recover unobserved productivity

$$\begin{aligned}\omega_{jt} = \; & \tfrac{\eta}{1+\eta}\Big[\delta_1 + [(1-\beta_l) - \tfrac{1}{\eta}\beta_l]l_{jt} + s_{jt} - p_{It} - \left(1+\tfrac{1}{\eta}\right)\beta_k k_{jt} \\ & + \left(\tfrac{1}{\eta}\right) q_{mt} + \left(\tfrac{1}{\eta}\right) x'_{jt}\beta_x\Big]\end{aligned} \tag{8}$$

where $\delta_1 = -ln(\beta_l) - ln(1+\frac{1}{\eta}) - \beta_0(1+\frac{1}{\eta}) - \ln E[e^{u^p_{jt}}] + \left(\frac{1}{\eta}\right)\ln E[e^{u^d_{jt}}] + \left(\frac{1}{\eta}\right)\ln E[e^{\upsilon_{jt}}]$. We then substitute the inverse demand function (8) in the value added generating function (6).[29]

[28] The average wage contains both price of labor and its composition, e.g., ages, gender, and skill groups. Our measure of wage is a good reflection of exogenous changes in the price of labor because the 22% growth in average retail wages during the period (Table 2) is in line with the 24% growth in aggregate real wages in Sweden.

[29] The condition for identification is that the variables in the parametric part of the model are not perfectly predictable (in the least square sense) by the variables in the non-parametric part (Robinson 1988). Including additional variables that affect productivity guarantees identification, i.e., there cannot be a functional relationship between the variables in the parametric and non-parametric parts (Newey et al. 1999). For example, z_{mt} cannot

It is important to stress again that we can estimate the value added generating function coefficients (6) because we have assumed that labor is a static variable. Comparing with non-parametric approaches, our estimator is more transparent how real wages and unobserved demand shocks affect labor demand. Ackerberg et al. (2006) (ACF) is an alternative estimator for which we show results in the empirical part. We use OLS and ACF estimators as benchmarks, i.e., without controlling for unobserved prices and local market characteristics. In ACF, labor has dynamic implications and labor demand is assumed to be a non-parametric function. Controlling for unobserved prices and local market conditions in ACF in a similar way, we expect that the elasticity of scale to increase. Dorazelski and Jamandreu (2009) discuss the relative merits of the parametric and non-parametric approaches.

■ **Selection.** Selection can be essential in retail markets because large stores are more likely to survive larger shocks to productivity than are small stores. Stores' decisions to exit in period t depend directly on ω_{jt}, and therefore the decision is correlated with the productivity shock ξ_{jt}. To estimate β_l and β_k while controlling for selection, we use predicted survival probabilities $\mathscr{P}_{t-1}$ (Appendix B gives a detailed description of selection). Substituting the survival probabilities and the inverse labor demand function (8) into (6) yields the final value added generating function that we estimate:

$$\begin{aligned} y_{jt} = \quad & \left(1+\frac{1}{\eta}\right)[\beta_0+\beta_l l_{jt}+\beta_k k_{jt}] - \frac{1}{\eta}q_{mt} - \frac{1}{\eta}x'_{jt}\beta_x + \left(1+\frac{1}{\eta}\right)h(\mathscr{P}_{t-1},\omega_{jt-1}) + \\ & \left(1+\frac{1}{\eta}\right)\xi_{jt} - \frac{1}{\eta}\upsilon_{jt} - \frac{1}{\eta}u^d_{jt} + \left(1+\frac{1}{\eta}\right)u^p_{jt}. \end{aligned} \tag{9}$$

■ **Estimation strategy.** The estimation of our extended Olley and Pakes model adjusted for retailers (EOP) consists of two parts. First, we use a probit model with a third order polynomial to estimate survival probabilities, which are then substituted into (9). Then, we estimate (9) using the sieve minimum distance procedure proposed by Ai and Chen (2003) and Newey and Powell (2003) for i.i.d. data. The goal is to obtain an estimable expression for the unknown parameters $\boldsymbol{\beta}$ and h_H, where H indicates all parameters in $h(\cdot)$. We approximate $h(\cdot)$ by a third order polynomial expansion in ω_{jt-1}, given by (8).[30] We use a tensor product polynomial series of labor (l_{jt-1}), capital (k_{jt-1}), total wages (s_{jt-1}), the consumer price index for food products (p_{It}), and local market conditions (z_{mt-1}) including large entrants and population density, plus local political preferences (pol_{mt}) as instruments. This set of instruments is also used to estimate the optimal weighting matrix. Using GMM, the parameters ($\boldsymbol{\beta}, h_H$) are then jointly estimated. Since there are non-linearities in the coefficients, we use the Nelder-Mead numerical optimization

be perfectly predicted from ω_{jt}.

[30] As a robustness check, we also expand $h(\cdot)$ using a fourth order polynomial, and the results are similar.

method to minimize the GMM objective function[31]

$$\min_{\boldsymbol{\beta},h_H}\left[\frac{1}{N}W'\psi(\cdot;\boldsymbol{\beta},h_H)\right]' A\left[\frac{1}{N}W'\psi(\cdot;\boldsymbol{\beta},h_H)\right] \tag{10}$$

where A is the weighting matrix defined as $A=\left[\frac{1}{N}W'\psi\psi'W\right]^{-1}$, W is the matrix of instruments, and $\psi_{jt}(\boldsymbol{\cdot};\boldsymbol{\beta},h_H)=\left[\left(1+\frac{1}{\eta}\right)\xi_{jt}-\frac{1}{\eta}\upsilon_{jt}-\frac{1}{\eta}u^d_{jt}+\left(1+\frac{1}{\eta}\right)u^p_{jt}\right]$. Estimation is done at the industry level, controlling for local conditions. Estimation results at county level (21 municipality groups) are available from authors. An advantage of estimating at county level is that we obtain the mark-ups at the county level. The major disadvantage is that we loose efficiency in estimation in the small counties.[32] Another advantage of using counties is that they are responsible for inter-municipality implementation of entry regulation. However, we control for municipality characteristics in the estimation. Appendix B presents a detailed description of the estimation procedure.

■ **Results: store TFP.** We estimated coefficients of the value added generating function using OLS, the Ackerberg et al. (2006) (ACF) two-stage estimator, and five specifications of our extended Olley and Pakes estimation. These five are: DP1 - productivity and persistent demand shocks follow the same AR(1) process, i.e., an updated version of the Blundell and Bond (2000) estimator; DP2 - productivity and persistent demand shocks follow different AR(1) processes; EOPs - productivity follows a nonlinear Markov process, and we control for selection, but not for prices or local market characteristics; EOPm - productivity follows a nonlinear Markov process, and we control for prices and local market characteristics but not for selection; and EOPms - productivity follows a nonlinear Markov process and we control for prices, selection and local market characteristics.[33] We include number of large entrants and population density, as local market covariates in the demand equation. We control for endogeneity of large entrants by using political preferences in local markets as instruments.[34]

[31]This simplex method converges quickly and is more robust to the starting values than quasi-Newton methods such as BFGS. Our EOP estimation procedure is written in R (www.r-project.org). The procedure is more computationally demanding when controlling for selection. However, estimation at the county level reduces computing time substantially.

[32]Another reason for estimating at county level is that our method requires more observations than is available at municipality level.

[33]To make our DP1 and DP2 estimators comparable with ACF and the dynamic panel estimator, we assume that productivity ω_{jt} is an AR(1) process, i.e., $\omega_{jt}=\rho_1\omega_{jt-1}+\xi_{jt}$. We use k_{jt} and l_{jt-1} as instruments, i.e., they are assumed to be uncorrelated with the shocks ξ_{jt} and υ_{jt}. However, we need an additional moment in DP to identify ρ_1 and therefore assume that the shock ξ_{jt} is uncorrelated with $(\omega_{jt-1}+u^d_{jt-1})$. In ACF, we use k_{jt} and l_{jt-1} as second stage instruments, i.e., labor is chosen with full knowledge of ω_{t-1}. Ackerberg et al. (2006) provide a detailed comparison between OP-type estimators and dynamic panel estimators.

[34]As noted earlier, we base on the political preferences in each municipality because no major policy reforms took place in Sweden during the study period, and we do not have access to the number of applications and rejections in the planning process. The Social Democrats are the largest party nationally with 40.6% of seats on average, collaborating with the Left Party (8%) and the Green Party (4.2%). The non-socialist group consists of the Moderate Party (18%), most often together with the Center Party (13.2%), Christian Democrats (5.9%), and the Liberals (5.6%). 22% of the municipalities had a non-socialist majority during 1996-1998, increasing to 32% during 1999-2002. The non-socialists had 8.6%-85%, averaging 40.7% (1996-1998) and 44.1% (1999-2002). The correlation between the non-socialist share of seats and the number of large entrants is 0.005, or 0.086 if

EOPs, EOPm, and EOPms require estimation of one non-parametric function, in contrast to ACF, which requires two. A major advantage of DP1, DP2, EOPm and EOPms is that they control for unobserved prices which otherwise might create a downward bias in the scale estimator (Klette and Griliches 1996). Another advantage is that the correction for omitted prices also yields an estimate of market output, which makes it possible to compute the elasticity of substitution η and an average industry mark-up.

Table 3 has two columns for each of the DP and EOP specifications. Column (1) shows the coefficients including elasticities, and Column (2) the larger true estimated coefficients, without elasticity. Since all specifications use deflated value added, we use Column (1) to compare OLS and ACF with DP and EOP.

The elasticity of scale estimate in the DP and EOP regressions is greater than in OLS (1.115) and ACF (0.931), it varies between 1.140 (EOPs) and 1.426 (DP2). The minimum point estimate of labor is 0.686 (DP2) and the maximum is 0.948 (OLS). The minimum point estimate of capital is 0.116 (EOPm) and the maximum is 0.426 (DP2). Controlling for local market characteristics is important: Including the number of large entrants and population density in the price equation change the demand elasticity and capital estimates substantially, making both smaller. When we allow productivity to follow an AR(1) process (DP1, DP2), estimates of capital are over 3 times larger than in EOP. The estimated productivity transition (ρ_1) is about 0.4 in both DP1 and DP2, i.e., a rather low persistency in over time. Furthermore, the estimated demand elasticities (-5.674 in DP1, -3.198 in DP2) seem unreasonably high in absolute value for retail food (Hall 1988). To test the assumption of linearity in productivity, we regress current productivity, recovered from DP1 and DP2, on a third order polynomial extension of previous productivity. The coefficients of ω_{jt-1}^2 and ω_{jt-1}^3 are statistically different from zero, indicating that productivity does not follow an AR(1) process. This might be one of the reasons for the large values of capital (over 0.4) in the DP specifications. We therefore recognize that it is important to allow for a nonlinear Markov process in productivity.

In ACF, EOPs, EOPm, and EOPms productivity follows a nonlinear Markov process. As noted, comparing with DP, the capital coefficients are smaller and the labor coefficients larger. As theory suggests the coefficients on both capital and labor decrease when controlling for prices.[35]

EOPs and EOPms (as well as ACF) control for selection. Theory and empirical investigations then predict a lower labor and higher capital coefficients (Ackerberg et al. 2007).[36] The capital coefficient in EOPms (0.145) is in fact larger than in EOPm (0.116), but smaller than

we exclude the Center Party, which is typically strong in the countryside where there is less likely to be large entrants. As we expect, we observe more large entrants in municipalities with non-socialist local government. We use a dummy for non-socialist majority in the estimations.

[35]If we do not control for unobserved demand shocks, we expect the coefficients of labor and capital to be upper biased. The reason is the positive correlation between inputs and demand shocks. In case that demand shocks are still present the coefficients would thus decrease even more. Nevertheless, our results when controlling for prices show that both coefficients decrease.

[36]Since stores with large capital stock can survive even if they have low productivity, we expect selection to induce a negative correlation between capital and the disturbance term in the selected sample.

in ACF (0.163). Controlling for selection in EOPms yields a smaller labor coefficient 0.840 than in EOPs (0.945). Those results are in line with the OP literature.

The coefficients on large entrants and population density are negative and statistically significant in all specifications. The lowest demand elasticity is (-2.96) is in EOPms, i.e., when we allow productivity to follow a nonlinear process and control for selection. Thus, the implicit assumption $\eta = -\infty$, often used in empirical studies, does not hold. In EOPms the the mark-up, defined as price over marginal cost, is 1.509. Our mark-up is consistent with previous findings based on retail data (see, e.g., Hall 1988).

Summarizing, it seems important to allow for nonlinearities in the productivity process and to control for prices, local market characteristics and selection (Ackerberg et al. (2006)). When estimating productivity in retail markets, it is thus central to deal with omitted price bias, unobserved demand characteristics and selection.

4 Large entrants and productivity

Next we proceed to investigate whether large entrants influence exit and productivity growth of surviving stores. Our goal is to evaluate whether large entrants have a greater impact on one part of the productivity distribution than another. To do this, we use TFP estimated by EOPms that allows for nonlinearities and selection, and DP2 that guarantees to clear out persistent demand shocks from productivity. Based on theories using dynamic models with heterogenous firms, our hypothesis for how increased competition from large entrants influences reallocation of resources is: Exit of low productive stores and/or higher productivity growth among surviving stores (Hopenhayn 1992, Melitz 2003, Syverson 2004, Asplund and Nocke 2006). We consider the role of large entrants for productivity levels, transitions in the productivity distribution in local markets, exit and productivity growth.[37] Finally, we decompose aggregate productivity growth of all entrants, exits and incumbents (due to data constraints we cannot measure the contribution of large entrants to aggregate productivity growth).

■ **Productivity levels.** Figure 1 shows kernel density estimates of TFP (estimated by EOPms) in markets the year of, and the year after, large entry. Though the differences are small, both the upper- and lower tails of the distribution are higher after large entry. However, productivity is to some extent lower in the middle of the distribution following large entry. Mean TFP is lower the year after entry (-0.476) than the year of entry (-0.361) and the standard deviation is larger (Table 4, panel A). Using t-test, the null of equal means is rejected at the 1% significance level. Using F-test, the null of equal standard deviations is also rejected at the 1% level.[38]

■ **Transitions in the productivity distribution.** To explore changes in productivity dis-

[37] We primarily focus on changes after large entry because several permanent reasons might explain differences between markets with and without large entrants.

[38] Defining entry markets as municipalities with at least one large entrant, mean TFP is smaller in markets with entrants (-0.330) than in markets without (-0.132) and the standard deviation is larger (Table 4, panel B). The null of equal means and equal standard deviations are both rejected and the significant at the 1% level.

tributions in local markets we classify incumbents into six percentile bins (p10, p10-25, p25-50, p50-75, p75-90, p90) each year, based on their productivity. Then we follow movements between percentile bins or exit over time.

Regardless of large entry, more stores increase their productivity in the bottom part of the distribution, while more stores decrease their productivity in the top (except p90). In markets with large entrants, more incumbents stay in the same percentile from one year to another, i.e., the diagonal shares are all higher (Table 5, panel A). The highest share of incumbents stay in p90 (48.75%). In markets without entrants, most incumbents stay in p50-75 (42%). We discuss productivity growth in detail in Section 4.2, but first we analyze exit.

4.1 Exit

Over 50% of the exits come from the two lowest percentiles (p10, p10-25) in markets with large entrants, but less than 42% in markets without large entrants (Table 5, panel A). Large entrants thus result in more exit among low productive stores. In markets without large entrants, more stores have lower productivity and yet continue to operate. While exit mainly occurs from the bottom part of the distribution, entrants are found across the whole distribution (not reported) as in previous findings in retail markets (Foster et al. 2006).

According to our model, the stopping rule implies that the decision to exit depends on productivity, capital stock and local market characteristics, i.e., large entrants and population density (see Section 3). Stores decide whether to exit or continue in the beginning of each period based on information regarding market conditions and we thus use large entrants in the previous year. Based on the stopping rule we show probit regressions of exit

$$Pr(exit_{jt} = 1 | e^L_{mt-1}, k_{jt-1}, p^{dens}_{mt-1}) = \phi(\gamma_0 + \gamma_e e^L_{mt-1} + D_{jt-1} * e^L_{mt-1}\Gamma + \gamma_k k_{jt-1} + \gamma_p p^{dens}_{mt-1} + \alpha_t) \tag{11}$$

where $exit_{jt}$ is equal to one if a store exit and otherwise zero; k_{jt} is log of capital; e^L_{mt-1} is the number of large entrants; $D_{jt-1} * e^L_{mt-1}$ are interaction terms between productivity percentile dummies and the number of large entrants; p^{dens}_{mt-1} is log of population density; ϕ is the cumulative distribution function of the standard normal; α_t is a vector of year dummies; and ε_{jt} is an i.i.d. error term; p^{dens}_{mt-1} and e^L_{mt-1} constitute z_{mt-1} in our model in Section 3.

Table 6 shows regression results for the probability of exit. The first specification (columns 1 and 3) relies on the pure stopping rule and includes productivity, capital, large entrants and population density. In line with both theory and previous empirical studies (Olley and Pakes 1996, Pavcnik 2002), exit is less likely if productivity and capital stock are high for both the nonlinear (column 1) and linear (column 3) productivity process. That is, stores with lower productivity and capital stock are more likely to exit. Moreover, exit is more common from markets if population density is high whereas the coefficient on large entrants is positive but not significant at conventional significance levels.

The expanded specification (columns 2 and 4) includes interaction terms of large entrants with the six productivity dummies, using the middle group (p50-75) as reference. In the non-linear estimation (column 2), the coefficient on large entrants is now negative and statistically significant at the 10 percent level. The coefficients on the interaction terms are all positive and jointly significant with the coefficient of large entry for p10 and p25-50. Exit is 0.29 percentage points more likely after large entry for stores is in the bottom part of the productivity distribution (p10 or p25-50) than for those in the middle. For the linear productivity process (column 4), the interaction terms are not significant, most likely because of lack of data (a store needs to be at least three years in the data).

To summarize, we find evidence that exit occurs from the bottom part of the productivity distribution after large entry which truncates the distribution from below in line with our hypothesis (Hopenhayn 1992, Olley and Pakes 1996, Melitz 2003, Syverson 2004, Asplund and Nocke 2006).

4.2 Productivity growth of incumbents

Productivity growth is given by the difference between expected productivity in time t and productivity in $t-1$: $\omega_{jt} - \omega_{jt-1}$. We only consider productivity growth of incumbents, and thus exclude stores that enter or exit. Figure 2 shows that incumbent productivity growth is higher the year after entry in markets with large entrants than in market without, with exception of the upper part. Mean productivity growth of incumbents (about 13%) and its standard deviation (about 0.7) are similar in markets with and without large entrants (Table 7, panel B). Neither the t-test nor the F-test can reject the null of equal mean values and standard deviations.

In markets with large entrants, Figure 3 shows a striking improvement in incumbent productivity growth between the year of, and the year after, entry: TFP growth is higher in all parts of the distribution after entry. Mean productivity growth of incumbents is -7.8% the year of large entry, whereas it is 12.4% the year after (Table 7, panel A). The t-test of equal mean values is rejected at the 1% significance level. The standard deviation is larger after entry, 0.75 compared to 0.66. Using the F-test, the null of equal standard deviations is also rejected at the 1% significance level.

Although Figure 2 and 3 indicate that large entrants might have an impact on productivity growth, we need to isolate the role of large entrants from store and market characteristics. Therefore, we regress the number of large entrants on incumbents' productivity growth the year after large entry,

$$\theta_{jt} = \alpha_0 + \alpha_e e^L_{mt-1} + D_{jt-1} * e^L_{mt-1}\boldsymbol{\alpha} + \alpha_p p^{dens}_{mt-1} + \alpha_m + \alpha_t + \varepsilon_{jt} \tag{12}$$

where $\theta_{jt} = \omega_{jt} - \omega_{jt-1}$ is incumbents' productivity growth between periods *t-1* and t; e^L_{mt-1} is the number of large entrants; $D_{jt-1} * e^L_{mt-1}$ are six interaction terms between percentile productivity dummies and large entrants; p^{dens}_{mt-1} is population density; α_t and α_m are vectors of time and market dummies; and ε_{jt} is an i.i.d. error term.

To isolate the impact of large entrants, we control for unobserved local market heterogeneity by using fixed effects for local markets and years. To control for endogeneity because large entry depends on the productivity of incumbents, we use different specifications of the one-step GMM estimator. Table 8 shows the regression results. GMM specification (1) uses lagged political preferences, lagged population density, and lagged income as instruments for large entrants; GMM specification (2) uses lagged large entrants (e^L_{mt-2}) plus lagged population density and lagged income as instruments. It is important to note that adding income as independent variable does not change our results. Since we get consistent results with all estimators, we primarily discuss the results of GMM (1) with TFP estimated by EOPms.

The coefficient on large entrants is positive and significant at the 1% level when we estimate with large entrants and productivity dummies but no interaction terms (Table 8). On average, large entrants thus increase productivity growth among incumbents. But large entrants have a greater impact on some parts on the incumbents' productivity distribution than others. The coefficient on large entry is then negative, whereas those on the interaction terms are all positive, and all significant at the 1% level. The coefficients of large entrants and the interaction terms are jointly significant. As a result of large entry, low productive incumbents increase their productivity growth, by 14% for those in p10 instead of in the middle group (p50-75), by 5% (for p10-25) and by 4% (for p25-50). On the other hand, large entry reduces productivity growth of incumbents in the upper distribution percentiles by -3% (for p75-90) and -7% (for p90), relative to the middle group. The growth increase is thus largest for surviving stores with low productivity, and then declines with survivors' productivity.[39] The results for the linear TFP process (DP2) are consistent with the ones we find for the nonlinear process (EOPms). The marginal effects of large entrants on productivity growth is larger for DP2 than for EOPms. One likely explanation is that DP2 only captures strong incumbents due to that stores need to be at least three years in the data. Our findings are in line with our hypothesis that competition increases productivity growth of incumbents.

The coefficient on population density is positive and significant at the 1% level, i.e., productivity growth is on average higher in dense markets, also in line with theoretical expectations (Syverson 2004).

■ **Decomposition of aggregate productivity growth.** Because of data constraints we can not decompose the contribution of large entrants to aggregate TFP growth but only the contribution of all entrants, exits, and incumbents. We use three recent decompositions, the one by Foster et al. (2001) (FHK), Griliches and Regev (1995) (GR), and Melitz and Polanec (2009) (MP) which is a dynamic version of the static decomposition by Olley and Pakes (1996). All

[39] In a previous version of the paper we investigate how large entrants affect the distribution of local market productivity, without controlling for large entry and unobserved demand shocks when estimated productivity. We found that productivity dispersion increases as a result of large entrants; the most productive incumbents become more productive, and the least productive become less productive. One explanation why the least productive stores become less productive is that a demand shock hit them after large entry, but they still find demand to survive. This indicates existence of demand shocks in productivity. Controlling for unobserved demand shocks when estimate productivity we find that the least productive become more productive. These results indicate importance of unobserved demand shocks when estimate productivity.

decompositions are discussed in detail in Appendix C. Aggregate TFP growth was 8% from 1997 to 2002 (Tables 9 and 10). In both FHK and GR, incumbents that continue for the whole period contribute to about 6%. Net entry stand for 1.85% in GR and 1.47% in FHK. Incumbents that increase both productivity and market shares stand for almost 1% of growth in FHK.

In MP, entrants and exits only have a positive contribution when their aggregate productivity is larger than that of continuing stores in the same period. As we expect, incumbents contribution is larger in MP than in GR and FHK (9.53%). Incumbents are more productive than both entrants and exits since entrants contribution is -3.67% and exits 2.14%. Among incumbents, those that obtain productivity improvements are central, whereas reallocation of market shares among them has a negative contribution. The direct effect of exits is the largest component showing that exits contribute to growth as they are less productive than incumbents. The indirect effect shows that the covariance between market shares and productivity is greater for entrants and exits than for incumbents. The decomposition results confirm our results based on large entrants, i.e., that incumbents that increase their productivity, and low productive stores that exit foster productivity growth.

5 Conclusions

The present study gives new insights into competition and productivity differences among retail stores. Net entry is found to foster almost all labor productivity growth in the U.S. retail sector (Foster et al. 2006). Multi-factor productivity in retail markets has however rarely been studied, contrary to manufacturing. We provide a first attempt to use recent advances in structural estimation of production functions to estimate total factor productivity in retail markets. Based on recent extensions of the Olley and Pakes' (1996) framework, we provide a model that takes key features of retail markets into account. In particular, we investigate one of the most crucial trends in retail markets: entry by large ("big-box") stores. On both sides of the Atlantic, the pros and cons of the big-box format have been widely debated (the Wal-Mart effect). We analyze whether large entrants force low productive stores out from the market and increase productivity growth among surviving stores with different positions in the productivity distribution. We use political preferences in local markets to control for endogeneity of large entrants. Our empirical application relies on detailed data on all retail food stores in Sweden, which is representative to many European markets in terms of market structure and regulation.

The results show that when estimating retail productivity, it is central to control for local market characteristics, and for selection, and to allow for nonlinearities in the productivity process. We recognize that large entrants clearly drive reallocation of resources towards more productive stores. After large entry, low productive stores exit. In addition, large entrants on average increase productivity growth of incumbent stores. The magnitude of the growth increase varies with an incumbent's position in the productivity distribution. The increase in growth declines with productivity. Growth thus increases relatively more among low productive survivors than among high productive ones. The productivity distribution thus gets truncated from below and dispersion decreases. We conclude that entry by big-box stores works as a catalyst for retail productivity growth.

Our findings contribute with knowledge to competition policy because entry regulation issues greatly concern policy makers in Europe, where such regulations are generally much more restrictive than in the U.S. As an example, the European Parliament recently highlighted an investigation of supermarket dominance (European Parliament 2008). We argue that a more restrictive design and application of entry regulations can hinder reallocation towards more productive units and thus hinder aggregate productivity growth. Note however that we clarify the indirect link between regulation, large entrants, and productivity because the numbers of approvals and rejections are not available. Besides productivity, entry regulations compound a wide range of other aspects. How to balance potential productivity growth against increased traffic and broader environmental issues are interesting issues for future research. It would also be interesting to apply our extended Olley and Pakes (1996) framework to other service markets such as banking and health care services. Future work would also benefit from using fully dynamic models (Dunne et al. 2005, Beresteanu and Ellickson 2006, Aguirregabiria et al. 2007, Holmes 2008) that would more carefully consider the importance of sunk costs, chain effects, and market adjustments.

References

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," *Handbook of Econometrics*, 6, 4171–4276.

Ackerberg, D., K. Caves, and G. Fraser (2006): "Structural Identification of Production Functions," Mimeo, UCLA.

Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl (2003): "Entry and Productivity Growth: Evidence from Micro-Level Panel Data," *Journal of the European Economic Association*, 2, 265–276.

——— (2009): "The Effects of Entry on Incumbent Innovation and Productivity," *Review of Economics and Statistics*, 91(1), 20–32.

Aguirregabiria, V., R. Mira, and H. Roman (2007): "An Estimable Dynamic Model of Entry, Exit and Growth in Oligopoly Retail Markets," *American Economic Review*, 97(2), 449–454.

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Asplund, M., and R. Friberg (2002): "Food Prices and Market Structure in Sweden," *Scandinavian Journal of Economics*, 104(4), 547–567.

Asplund, M., and V. Nocke (2006): "Firm Turnover in Imperfectly Competitive Markets," *Review of Economic Studies*, 73, 295–327.

Baily, M., C. Hulten, and D. Campbell (1992): "Productivity Dynamics in Manufacturing Establishments," *Brooklyns Papers on Economic Activity*, pp. 187–249.

Bartelsman, E., and M. Doms (2000): "Understanding Productivity: Lessons from Longitudinal Microdata," *Journal of Economic Literature*, 38(3), 569–594.

Basker, E. (2005): "Job Creation or Destruction? Labor-Market Effects of Wal-Mart Expansion," *Review of Economics and Statistics*, 87(1), 174–183.

——— (2007): "The Causes and Consequences of Wal-Mart's Growth," *Journal of Economic Perspectives*, 21(3), 177–198.

Basker, E., and M. Noel (2009): "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry," *Journal of Economics and Management Strategy*, 18(4), 177–198.

Beresteanu, A., and P. Ellickson (2006): "The Dynamics of Retail Oligopoly," Mimeo, Duke University.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Bertrand, M., and F. Kramarz (2002): "Does Entry Regulation Hinder Job Creation? Evidence from the French retail industry," *Quarterly Journal of Economics*, 117(4), 1369–1413.

Blundell, R., and S. Bond (2000): "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 31(3), 549–569.

Blundell, R., R. Griffith, and J. Van Reenen (1999): "Market Share, Market Value, and Innovation in a Panel of British Manufacturing Firms," *Review of Economic Studies*, 66(3), 529–554.

Bond, S., and M. Söderbom (2005): "Adjustment Costs and the Identification of Cobb-Douglas Production Functions," Working Paper 05/04, Institute of Fiscal Studies.

Boone, J. (2000): "Competitive Pressure: The Effects on Investments in Product and Process Innovation," *RAND Journal of Economics*, 31(3), 549–569.

Boylaud, O., and G. Nicoletti (2001): "Regulatory reform in retail distribution," OECD Working Paper 32.

Buettner, T. (2004): "R&D and the Dynamics of Productivity," Mimeo, London School of Economics.

Caves, R. (1998): "Industrial Organization and New Findings on the Turnover and Mobility of Firms," *Journal of Economic Literature*, 36(4), 1947–1982.

Chen, X., and S. Ludvigson (2007): "Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models," Mimeo, New York University.

De Loecker, J. (2009): "Product Differentiation, Multi-Product Firms and Structural Estimation of Productivity," Mimeo, New York University.

Doraszelski, U., and J. Jaumandreu (2009): "R&D and Productivity: Estimating Production Functions when Productivity is Endogenous," Mimeo, Harvard University.

Dunne, T., S. Klimek, M. Roberts, and Y. Xu (2005): "Entry and Exit in Geographical Markets," Mimeo, Pennsylvania State Univeristy.

Ericson, R., and A. Pakes (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–83.

European Parliament (2008): "On investigating and remedying the abuse of power by large supermarkets operating in the European Union," Written Declaration 0088, European Parliament.

Fishman, C. (2006): *The Wal-Mart Effect*. Penguin Press, New York.

Foster, L., J. Haltiwanger, and C. Krizan (2001): "The Link Between Aggregate and Micro Productivity growth: Evidence from Retail Trade," NBER Working Paper 13783.

——— (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.

Griffith, R., and H. Harmgart (2005): "Retail Productivity," *International Review of Retail, Distribution and Consumer Services*, 15(3), 281–290.

Griliches, Z., and H. Regev (1995): "Productivity and Firm Turnover in Israeli Industry: 1979-1988," *Journal of Econometrics*, 65, 175–203.

Hall, R. (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96(5), 921–947.

Hicks, M. J. (2007): *The Local Economic Impact of Wal-Mart*. Cambria Press, Youngstown, New York.

Hoj, J., T. Kato, and D. Pilat (1995): "Deregulation and privatisation in the service sector," OECD Working Paper 25.

Holmes, T. (2008): "The Diffusion of Wal-Mart and Economies of Density," NBER Working Paper 13783.

Hopenhayn, H. (1992): "Entry, Exit and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60(5), 1127–1160.

Hulten, C., and F. Wykoff (1981): chap. The measurement of Economic Depreciation, in C. Hulten (ed.), *Depreciation, Inflation, and Taxation of Income from Capital*, Urban Institute Press, Washington, D.C.

Jia, P. (2008): "What Happens when Wal-Mart comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76(6), 1263–1316.

Jovanovic, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50(5), 649–670.

Katayama, H., S. Lu, and J. Tybout (2003): "Why Plant-Level Productivity Studies are Often Misleading, and an Alternative Approach to Inference," NBER Working Paper 9617.

Klette, T., and Z. Griliches (1996): "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11(4), 343–361.

Levinsohn, J., and M. Melitz (2002): "Productivity in a Differentiated Products Market Equilibrium," Mimeo, Harvard University.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–341.

MacDonald, J. (1994): "Does Import Competition Force Efficient Production?," *Review of Economics and Statistics*, 76(4), 721–727.

Melitz, M. (2000): "Estimating Firm Level Productivity in Differentiated Product Industries," Mimeo, Harvard University.

——— (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695–1725.

Melitz, M., and S. Polanec (2009): "Dynamic Olley-Pakes Decomposition with Entry and Exit," Mimeo, Princeton University.

Newey, W., and J. Powell (2003): "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71(3), 1557–1569.

Newey, W., J. Powell, and F. Vella (1999): "Nonparametrric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67(3), 565–603.

Nickell, S. (1996): "Competition and Corporate Performance," *Journal of Political Economy*, 104(4), 724–746.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

Pavcnik, N. (2002): "Trade Liberalization Exit and Productivity Improvements: Evidence from Chilean Plants," *Review of Economic Studies*, 69(1), 245–276.

Pilat, D. (1997): "Regulation and Performance in the Distribution Sector," OECD Working Papers 180.

——— (2005): "Assessing the Productivity of the UK Retail Sector: Some Further Reflections," *The International Review of Retail, Distribution and Consumer Research*, 15(3), 291–296.

Porter, M. (1990): *The Competitive Advantage of Nations*. Macmillan Press.

Raith, M. (2003): "Competition, Risk, and Managerial Incentives," *American Economic Review*, 93, 1425–1436.

Reynolds, J., E. Howard, D. Dragun, B. Rosewell, and P. Ormerod (2005): "Assessing the Productivity of the UK Retail Sector," *International Review of Retail, Distribution and Consumer Research*, 15(3), 237–280.

Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Sadun, R. (2008): "Does Planning Regulation Protect Independent Retailers?," Center for Economic Performance Discussion Paper No 888, London School of Economics.

Schmidt, K. (1997): "Managerial Incentives and Product Market Competition," *Review of Economic Studies*, 64, 191–213.

Sivadasan, J. (2004): "Barriers to Entry and Productivity: Micro-Evidence from Indian Manufacturing Sector Reforms," Mimeo, University of Chicago.

Statistics Sweden (2005): "Hushållens ekonomi - vem lånar, vem konsumerar? (The Economy of Households: Who lends, who comsumes?)," Mimeo, Stockholm.

Swedish Competition Authority (2001:4): "Kan Kommunerna Pressa Matpriserna? (Can the Municipalities put Pressure on Prices?)," Technical Report 4, Stockholm.

——— (2004:2): "Konsumenterna, Matpriserna och Konkurrensen (Consumers, Retail Food Prices, and Competition)," Technical Report 2, Stockholm.

Swedish National Board of Housing, Building, and Planning (2005): "Shopping martters! Retailing and sustainable urban development," Technical Report.

Syverson, C. (2004): "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(2), 1181–1222.

Table 1: Characteristics of the Swedish Retail Food Market

A. DELFI						
Year	No. of stores	Large stores	Large entry	Mean sales space (m^2)	Total sales space (m^2)	Total sales
1996	4,664	905	21	538	2,510,028	129,326,000
1997	4,518	925	8	550	2,483,248	126,732,397
1998	4,351	926	9	587	2,552,794	130,109,604
1999	4,196	936	18	604	2,514,367	133,156,023
2000	3,994	948	23	654	2,587,952	138,314,044
2001	3,656	942	28	689	2,471,510	139,352,920
2002	3,585	932	5	718	2,525,084	142,532,944

B. FS-RAMS					
Year	No. of “multi-stores”	No. of employees	Total wages	Value added	Total sales
1996	3,714	74,100	9,882,234	18,319,407	141,743,876
1997	3,592	73,636	10,322,136	18,838,130	142,840,611
1998	3,482	74,696	10,766,043	19,185,120	147,726,647
1999	3,398	74,758	11,110,785	19,570,472	152,160,949
2000	3,287	77,180	11,536,063	20,389,492	154,106,865
2001	3,094	76,905	11,522,482	20,748,902	158,512,132
2002	3,067	80,931	12,081,931	22,473,696	179,335,162

NOTE: DELFI is provided by Delfi Marknadspartner AB and contains all retail food stores based on their geographical location (address). FS-RAMS is provided by Statistics Sweden and consists of all organization numbers in SNI-code 52.1, i.e., “multi-store” units that contain one store or several (e.g., due to the same owner). Sales (incl. 12% VAT), value-added, and wages are measured in thousands of 1996 SEK (1USD=6.71SEK, 1EUR=8.63 SEK). Sales in DELFI are collected by surveys and reported in classes, while sales are based on tax reporting in FS-RAMS. Therefore, total sales are lower in DELFI than in FS-RAMS. From 1996 to 2002, the total population in Sweden increased from 8,844,499 to 8,940,788.

Table 2: Medians of local market characteristics

Year	1997	1998	1999	2000	2001	2002
A. Markets with large entrants						
No. of stores	37.00	54.00	29.00	32.00	33.00	22.00
No. of all entrants	2.00	2.00	3.00	2.00	1.00	2.00
No. of all exits	3.00	2.00	2.00	3.00	1.00	-.-
Population	57,441.00	60,429.00	37,195.00	48,250.00	58,361.00	22,907.00
Population density	80.88	57,92.00	68.03	79.38	77.29	52.77
Per capita income	149.10	157.60	161.60	170.30	179.10	177.60
Store concentration (C_4)	0.53	0.49	0.62	0.60	0.53	0.70
Total no. of markets	10	9	20	20	23	6
B. Markets without large entrants						
No. of stores	15.00	15.00	15.00	14.00	13.00	14.00
No. of all entrants	0.00	0.00	1.00	0.00	0.00	0.00
No. of all exits	0.00	1.00	1.00	1.00	0.00	-.-
Population	14,827.00	15,133.00	14,322.00	14,154.00	14,068.00	15,207.00
Population density	25.80	25.78	25.22	25.60	24.75	26.20
Per capita income	143.30	149.10	155.90	162.50	168.40	175.90
Store concentration (C_4)	0.71	0.71	0.72	0.73	0.75	0.76
Total no. of markets	278	279	269	269	266	284

NOTE: 1996 is left out because entrants are not observed. Municipalities, considered as local markets, increase from 288 to 290 due to three municipality break-ups during the period. Stores, entrants and exits come from DELFI. Population density is defined as total population per square kilometer in the municipality. Concentrations (C_4) show the market share captured by the top four stores.

Table 3: Value added generating function estimates

	OLS	ACF	DP1		DP2		EOPs		EOPm		EOPms	
			(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Log no. of labor	0.948	0.768	0.754	0.916	0.686	0.900	0.845	0.921	0.945	1.205	0.840	1.269
	(0.005)	(0.057)	(0.001)		(0.001)		(0.001)		(0.001)		(0.007)	
Log of capital	0.167	0.163	0.400	0.485	0.426	0.400	0.212	0.232	0.116	0.147	0.145	0.201
	(0.003)	(0.050)	(0.001)		(0.001)		(0.001)		(0.003)		(0.001)	
Market output $\left(-\frac{1}{\eta}\right)$			0.176		0.000		0.082		0.216		0.338	
			(0.002)		(0.001)		(0.001)		(0.001)		(0.001)	
Number of large entrants			-0.945	-5.371	-0.030	-0.984			-0.093	-0.430	0.034	0.100
			(0.002)		(0.000)				(0.001)		(0.004)	
Population density			-0.103	-0.421	-0.165	-0.529			-0.054	-0.252	-0.033	-0.186
			(0.002)		(0.000)				(0.001)		(0.001)	
Productivity transition (ρ_1)			0.417		0.449							
			(0.002)		(0.001)							
Demand shock transition (ρ_2)					0.353							
					(.)							
Scale ($\beta_l + \beta_k$)	1.115	0.931	1.402		1.426		1.140		1.362		1.420	
Demand elasticity (η)			-5.674		-3.198		-12.192		-4.659		-2.962	
Mark-up $\left(\frac{\eta}{1+\eta}\right)$			1.214		1.089		1.089		1.275		1.509	
Sargan (p-value)		0.081			0.134		0.134		0.172		0.152	
No. of obs.	23,521	16,186	15,640		15,640		15,640		15,640		15,640	

NOTE: The dependent variable is log of deflated value added. Labor is measured as number of full-time adjusted employees. All regressions include year dummies. OLS is ordinary least square regression. ACF is Ackerberg, Caves, and Fraser's (2006) two-stage estimation method; DP1 is linear estimation of equation (6) when ω_{jt} and υ_{jt} follow the same AR(1) process; DP2 is linear estimation of equation (6) when ω_{jt} and υ_{jt} follow two different AR(1) processes; EOPs is the semi-parametric estimation of equation (9) without local market characteristics but controlling for selection; EOPm is the semi-parametric estimation of equation (9) with control for prices and local market characteristics but not for selection; EOPms is the semi-parametric estimation of equation (9) specified in Section 3, i.e., we control for prices, local market characteristics, and selection. In the EOP specifications, columns (1) show estimated coefficients including elasticity (see Equation 6); columns (2) show estimated coefficients without elasticity. Reported standard errors (in parentheses) are robust to heteroscedasticity. In ACF, current capital stock and previous labor are used as instruments, and standard errors are computed using bootstrap. In EOP, two-step GMM is used for estimation. Market output is measured as the market share weighted output in the municipality. Demand refers to the elasticity of substitution. Mark-up is defined as price over marginal cost.

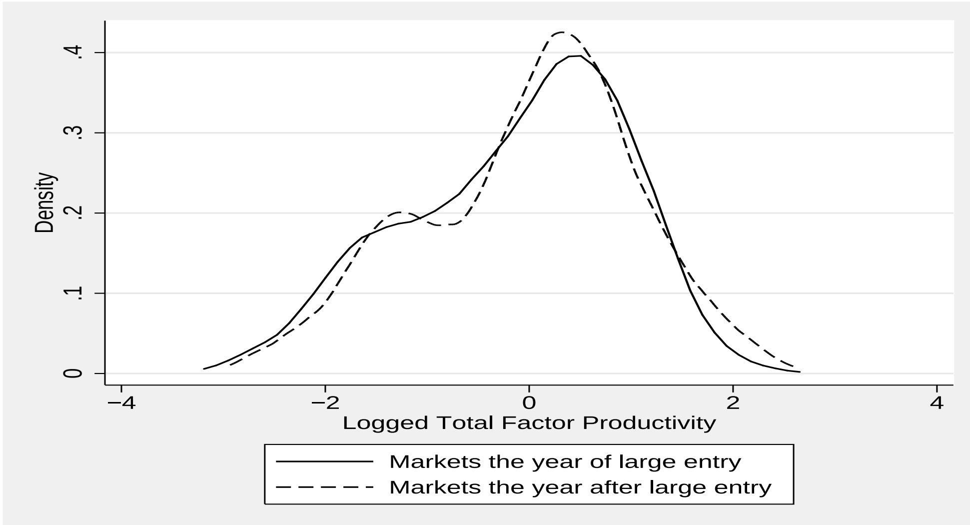

Figure 1: TFP kernel density estimates, incumbent stores in markets the year of, and the year after, large entry

Table 4: TFP and large entrants

A. Markets with large entrants	Mean	Std. Dev.
Year of entry	-0.361	1.286
Year after entry	-0.476	1.400
Test (p-value)	0.004	0.001
B. All markets		
With entry	-0.330	1.190
Without entry	-0.132	1.020
Test (p-value)	0.001	0.001

NOTE: This table summarizes TFP levels in markets before and after large entrants, and in markets with and without large entrants. T-test is used for mean, and F-test is used for standard deviation (p-values reported). TFP is estimated using the semi-parametric EOPms method described in Section 3. Large entrants are defined as the five largest store types in the DELFI data (hypermarkets, department stores, large supermarkets, large grocery stores, and other stores).

Table 5: Transition matrix from t-1 (column) to t (row)

Percentile	<p10	p10-p25	p25-p50	p50-p75	p75-p90	>p90	Exit
A. Markets with large entrants in t-1							
<p10	35.59	22.88	10.17	1.69	1.69	0.85	27.12
p10-p25	14.05	33.47	23.55	4.13	1.65	0.00	23.14
p25-p50	2.53	12.64	42.30	19.77	4.14	1.38	17.24
p50-p75	0.85	2.34	21.44	44.37	13.80	2.97	14.23
p75-p90	0.00	1.77	6.38	26.24	37.23	13.48	14.89
>p90	0.63	1.25	2.50	11.88	20.00	48.75	15.00
B. Markets without large entrants in t-1							
<p10	31.84	26.02	12.55	4.08	0.61	1.53	23.37
p10-p25	15.32	34.75	23.92	5.63	1.39	0.82	18.16
p25-p50	4.11	13.72	40.75	20.17	3.97	1.77	15.52
p50-p75	0.86	3.15	20.52	42.06	16.04	5.50	11.87
p75-p90	0.39	1.34	8.52	25.83	33.05	16.81	14.06
>p90	1.09	0.84	5.13	15.63	25.55	35.38	16.39

NOTE: TFP is estimated using the semi-parametric EOPms method described in Section 3. Municipalities are considered as local markets. Large entrants in period t-1 are defined as the five largest store types in the DELFI data (hypermarkets, department stores, large supermarkets, large grocery stores, and other stores).

Table 6: Regression results: Exit

	TFP nonlinear (EOPms)		TFP linear (DP2)	
	(1)	(2)	(3)	(4)
Log of productivity	-0.058		-0.013	
	(0.015)		(0.007)	
Large entrants$_{t\text{-}1}$	0.022	-0.152	0.014	0.061
	(0.037)	(0.088)	(0.057)	(0.109)
p10*Large entrants$_{t\text{-}1}$		0.293		0.121
		(0.118)		(0.155)
p10-p25*Large entrants$_{t\text{-}1}$		0.190		0.128
		(0.126)		(0.167)
p25-p50*Large entrants$_{t\text{-}1}$		0.293		-0.216
		(0.113)		(0.160)
p75-p90*Large entrants$_{t\text{-}1}$		0.071		-0.275
		(0.117)		(0.197)
p90*Large entrants$_{t\text{-}1}$		0.209		-0.036
		(0.129)		(0.192)
Log of capital	-0.083	-0.087	-0.077	-0.070
	(0.009)	(0.009)	(0.014)	(0.014)
Log of population density$_t$	0.018	0.020	-0.004	-0.004
	(0.008)	(0.008)	(0.011)	(0.011)
Year dummies	yes	yes	yes	yes
No. of obs.	11,132	11,132	7,376	7,376

NOTE: This table shows probit regressions on exit. TFP is estimated using the semi-parametric EOP method described in Section 3 (EOPms) and linear panel specification (DP2). Reported standard errors (in parentheses) are robust to heteroscedasticity. Large entrants in period t-1 are defined as the five largest store types in the DELFI data (hypermarkets, department stores, large supermarkets, large grocery stores, and other stores). We use six percentile bins for productivity in each market and year, with p50-75 used as reference group.

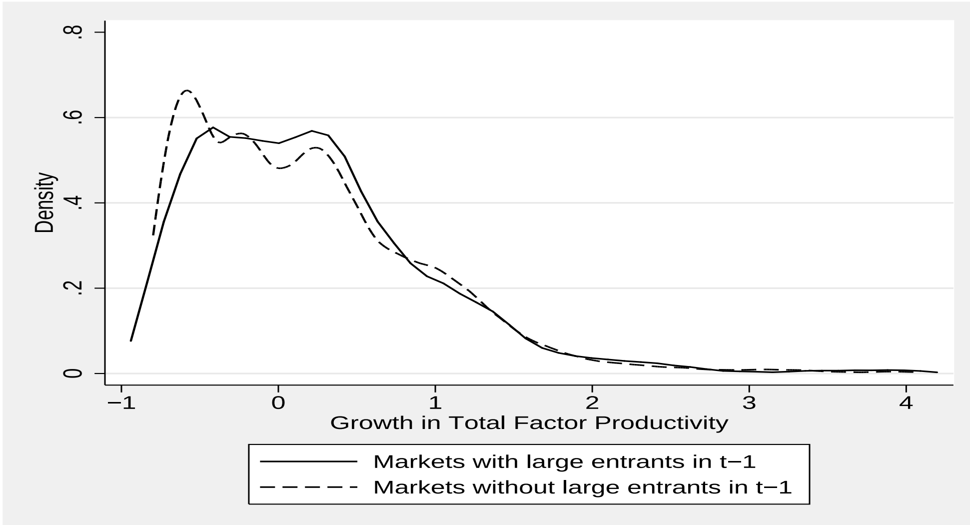

Figure 2: TFP growth kernel density estimates, incumbent stores in markets with and without large entrants

Table 7: TFP Growth and large entrants

A. Markets with large entrants	Mean	Std. Dev.
Year of entry	-0.079	0.664
Year after entry	0.124	0.752
Test (p-value)	0.001	0.001
B. All markets		
With entry	0.330	1.190
Without entry	0.132	1.020
Test (p-value)	0.997	0.349

NOTE: This table summarizes TFP growth in markets before and after large entrants, and in markets with and without large entrants. T-test is used for mean, and F-test is used for standard deviation (p-values reported). TFP is estimated using the semi-parametric EOPms method described in Section 3. Large entrants are defined as the five largest store types in the DELFI data (hypermarkets, department stores, large supermarkets, large grocery stores, and other stores).

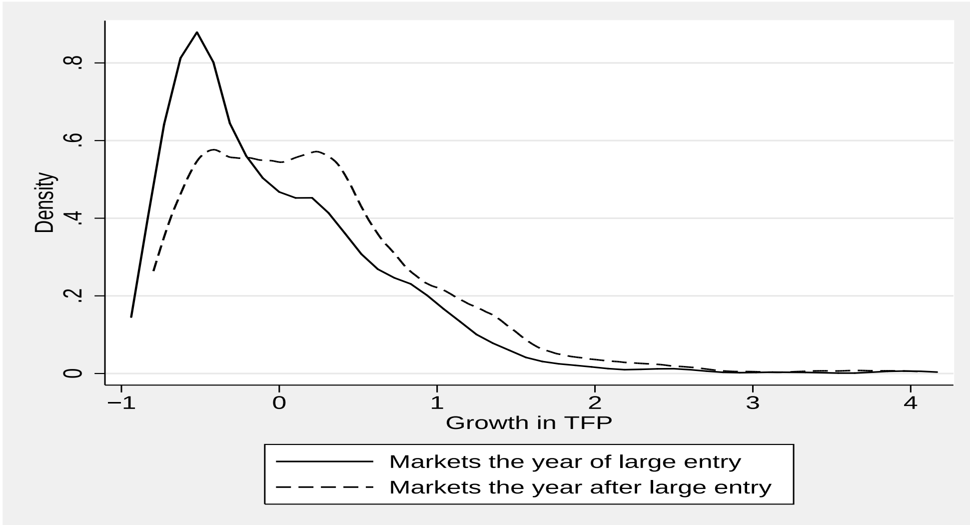

Figure 3: TFP growth kernel density estimates, incumbent stores in markets the year of, and the year after, large entrants

Table 8: Regression results: TFP growth

	TFP nonlinear (EOPms)					TFP linear (DP2)				
	OLS		Within	GMM		OLS		Within	GMM	
	(1)	(2)		(1)	(2)	(1)	(2)		(1)	(2)
Large entrants$_{t\text{-}1}$	0.040	-0.009	-0.021	-0.227	0.451	0.084	0.006	-0.073	-1.001	0.252
	(0.008)	(0.014)	(0.017)	(0.004)	(0.001)	(0.023)	(0.045)	(0.058)	(0.001)	(0.001)
p10* Large entrants$_{t\text{-}1}$		0.161	0.188	0.371	0.343		0.275	0.372	1.241	1.400
		(0.022)	(0.028)	(0.001)	(0.001)		(0.063)	(0.081)	(0.001)	(0.001)
p10-p25* Large entrants$_{t\text{-}1}$		0.079	0.107	0.281	0.593		0.154	0.273	1.087	1.840
		(0.020)	(0.025)	(0.001)	(0.001)		(0.060)	(0.081)	(0.001)	(0.001)
p25-p50* Large entrants$_{t\text{-}1}$		0.071	0.096	0.263	0.422		0.127	0.235	1.019	1.301
		(0.021)	(0.024)	(0.001)	(0.001)		(0.055)	(0.074)	(0.001)	(0.001)
p75-p90*Large entrants$_{t\text{-}1}$		0.004	0.017	0.196	-0.462		-0.084	0.005	0.798	-1.618
		(0.020)	(0.024)	(0.001)	(0.001)		(0.065)	(0.093)	(0.001)	(0.001)
p90*Large entrants$_{t\text{-}1}$		-0.031	-0.022	0.161	-0.314		-0.245	-0.090	0.625	-1.196
		(0.023)	(0.030)	(0.001)	(0.001)		(0.070)	(0.101)	(0.001)	(0.001)
Log of population density$_t$	0.180	0.110	-0.038	0.151	-1.166	-0.113	-0.264	-0.031	-0.244	-5.206
	(0.164)	(0.163)	(0.015)	(0.001)	(0.001)	(0.505)	(0.503)	(0.052)	(0.003)	(0.004)
Year fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Market fixed effects	yes	yes	yes	no	yes	yes	yes	no	yes	yes
Adjusted R^2	0.745	0.746				0.253	0.258			
J-test (p-value)				0.994	0.998				0.997	0.998
No. of obs.	13,626	13,626	13,626	13,626	9,774	7,376	7,376	7,376	7,376	4,884

NOTE: TFP is estimated using the semi-parametric EOP method described in Section 3 (EOPms) and linear panel specification (DP2). Standard errors reported in parentheses, and one-step GMM estimator is used. GMM (1) uses lagged political preferences as instruments for large entry, GMM (2) also adds lagged large entrants (in t-2), lagged population density, and lagged income. J-test refers to the test for overidentified restrictions in GMM models. Large entrants in period t-1 are defined as the five largest store types in the DELFI data (hypermarkets, department stores, large supermarkets, large grocery stores, and other stores). We use six percentile bins for productivity in each market and year, with p50-75 used as reference group.

Table 9: Decomposition of retail food productivity growth, 1997 to 2001

	Percentage of growth from					
Overall industry growth	Within stores (1)	Between stores (2)	Cross stores (3)	Entry (4)	Exit (5)	Net entry (4) - (5)
A. Baily, Hulten and Campbell (1992) / Foster, Haltiwanger and Krizan (2001)						
0.08	0.0572	0.0006	0.0075	0.0025	-0.0121	0.0147
B. Griliches and Regev (1995)						
0.08	0.0609	0.0005		0.0294	-0.0109	0.0185

NOTE: Decomposition using Equation (18) in Section 3; TFP is estimated using the semi-parametric estimation (EOPms) described in Section 3. Shares of local market sales are used as weights. Appendix C explains the decompositions in detail.

Table 10: Dynamic Olley and Pakes decomposition of TFP growth 1997-2002: Melitz and Polanec 2009

	Percentage of growth from					
	Surviving		Entrants		Exits	
Overall Industry Growth	Unweigh.	Cov	Unweigh.	Weigh.	Unweigh.	Weigh.
0.08	0.1124	-0.0171	-0.0893	-0.0367	0.117	0.0214

NOTE: TFP is estimated using the semi-parametric estimation EOP described in Section 3 and 4. Shares of local market sales are used as weights. Appendix C describes the decomposition in detail.

Appendix A: PBA and data sources

■ **Entry regulation (PBA).** On July 1, 1987, a new regulation was imposed in Sweden, the Plan and Building Act (PBA). Compared to the previous legislation, the decision process was decentralized, giving local governments power over entry in their municipality and citizens a right to appeal the decisions. Since 1987, only minor changes have been implemented in PBA. From April 1, 1992 to December 31, 1996, the regulation was slightly different, making explicit that the use of buildings should not counteract efficient competition. Since 1997, PBA has been more or less the same as prior to 1992. Long time lags in the planning process make it impossible to directly evaluate the impact of decisions. In practice, differences because of the policy change seem small (Swedish Competition Authority 2001:4). Nevertheless, PBA is claimed to be one of the major entry barriers, resulting in different outcomes, e.g., price levels, across municipalities (Swedish Competition Authority 2001:4, Swedish Competition Authority 2004:2). Municipalities are then, through the regulation, able to put pressure on prices. Those that constrain entry have less sales per capita, while those where large and discount stores have a higher market share also have lower prices.

■ **The DELFI data.** DELFI Marknadspartner AB collects daily data on retail food stores from a variety of channels: (1) public registers, the trade press, and daily press; (2) the Swedish retailers association (SSLF); (3) Kuponginlösen AB (which deals with rebate coupons collected by local stores); (4) the chains' headquarters; (5) matching customer registers from suppliers; (6) telephone interviews; (7) yearly surveys; and (8) the Swedish Retail Institute (HUI). Location, store type, owner, and chain affiliation are double-checked in corporate annual reports.

Each store has an identification number linked to its geographical location (address). The twelve store types, based on size, location, product assortment, etc., are hypermarkets, department stores, large supermarkets, large grocery stores, other stores, small supermarkets, small grocery stores, convenience stores, gas-station stores, mini markets, seasonal stores, and stores under construction.

Sales and sales space are collected via yearly surveys. Revenues (including VAT) are recorded in 19 classes. Due to the survey collection, a number of missing values are substituted with the median of other stores of the same type in the same local market. In total, 702 stores have missing sales: 508 in 1996, and 194 in later years. For sales space, all 5,013 values are missing for 1996, and are therefore replaced with the mean of each stores' 1995 and 1997 values. In addition, 2,810 missing sales space values for later years are replaced similarly. In total, 698 observations are missing both sales and sales space.

■ **The FS-RAMS data.** FS-RAMS contains all registered organization numbers in the different Swedish industries from 1996 to 2002. *Value added* is defined as total shipments, adjusted for inventory changes, minus costs of materials. *Labor* is the total number of employees. We deflated sales, value added, wages, and investment by the consumer price index (CPI) from IMF-CDROM 2005.

Capital is constructed using a perpetual inventory method, $k_{t+1}(1-\delta)k_t + i_t$. Since the data distinguishes between buildings and equipment, all calculations of the capital stock are done separately for buildings and equipment. In the paper, we include equipment in the capital stock. Including both equipment and buildings in the capital stock does not change the results, however. As suggested by Hulten and Wykoff (1981), buildings are depreciated at a rate of 0.0361, and equipment at 0.1179. In order to construct capital series using the perpetual inventory method, an initial capital stock is needed. We set initial capital stock to its first occurrence in FS-RAMS, defining entry as the first year in FS (some of the stores have been in FS since 1973).

Appendix B: Selection and estimation strategy

■ **Selection.** A store's decision to exit in period t depends directly on productivity ω_{jt}, so that the decision will be correlated with the productivity shock ε_{jt}. To identify β_l and β_k, we use estimates of survival probabilities, given by

$$\begin{aligned} Pr(\chi_t = 1|\underline{\omega}_t(k_t, z_{t-1}), \mathcal{F}_{t-1}) = \; & Pr(\omega_t \geq \underline{\omega}_t(k_t, z_{t-1})| \\ & \underline{\omega}_t(k_t, z_{t-1}), \omega_{t-1}) \\ = \; & P_{t-1}(i_{t-1}, l_{t-1}, k_{t-1}, s_{t-1}, p_{mt-1}, q_{mt-1}, \\ & x_{t-1}) \\ \equiv \; & \mathscr{P}_{t-1} \end{aligned} \tag{13}$$

where the second equality follows from (8). Controlling for selection, we can express the non-parametric function $h(\cdot)$ (the approximation of the conditional expectation $E[\omega_{jt}|\mathcal{F}_{t-1}]$) as a function of threshold market productivity $\underline{\omega}_t$ and the information set $\mathcal{F}_{t-1}$. As a result, threshold market productivity can be written as a function of $\mathscr{P}_{t-1}$ and $\mathcal{F}_{t-1}$. Substituting Equations (8) and (13) into (6) yields

$$\begin{aligned} y_{jt} = \; & \left(1+\tfrac{1}{\eta}\right)[\beta_0 + \beta_l l_{jt} + \beta_k k_{jt}] - \tfrac{1}{\eta} q_{mt} - \tfrac{1}{\eta} x_{jt-1})' \beta_x + h\left(\mathscr{P}_{t-1}, \delta_1 + \left[(1-\beta_l) - \tfrac{1}{\eta}\beta_l\right] l_{jt-1} \right. \\ & \left. -(1+\tfrac{1}{\eta})\beta_k k_{jt-1} + s_{jt-1} - p_{It-1} + \tfrac{1}{\eta} q_{mt-1} + \tfrac{1}{\eta} x_{jt-1})' \beta_x \right) + \left(1+\tfrac{1}{\eta}\right)\xi_{jt} - \tfrac{1}{\eta}\upsilon_{jt} \\ & -\tfrac{1}{\eta} u^d_{jt} + \left(1+\tfrac{1}{\eta}\right) u^p_{jt}. \end{aligned} \tag{14}$$

■ **Estimation strategy.** We first use a probit model with a third order polynomial to estimate the survival probabilities in (13). The predicted survival probabilities are then substituted into (9), which is estimated in the second step. We now turn to details about the estimation procedure of the latter step. The semi-parametric regression (9) is estimated using the sieve minimum distance (SMD) procedure proposed in Newey and Powell (2003) and Ai and Chen (2003) for i.i.d. data.[40] The goal is to obtain an estimable expression for the unknown parameter of interest, $\boldsymbol{\alpha} = (\boldsymbol{\beta}, h)'$. We denote the true value of the parameters with the subscript "a", so

[40]Chen and Ludvigson (2007) show that the SMD procedure and its large sample properties can be extended to stationary ergotic time series data.

that $\boldsymbol{\alpha}_a = (\boldsymbol{\beta}_a, h_a)^{'}$. The moment conditions could then be written more compactly as

$$E[\psi_{jt}(\boldsymbol{X_{jt}}, \boldsymbol{\beta}_a, h_a)|\mathcal{F}_t^*] = 0 \quad j = 1, \cdots, N \quad t = 1, \cdots, T \tag{15}$$

where N is the total number of stores, $\mathcal{F}_t^*$ is the information set at time t, and $\psi_{jt}(\cdot)$ is defined as

$$\begin{aligned} \psi_{jt}(\boldsymbol{X}_{jt}, \boldsymbol{\beta}_a, h_a) &\equiv \left[\left(1 + \tfrac{1}{\eta}\right)\xi_{jt} - \tfrac{1}{\eta}\upsilon_{jt} - \tfrac{1}{\eta}u_{jt}^d + \left(1 + \tfrac{1}{\eta}\right)u_{jt}^p\right] \\ &= y_{jt} - \left(1 + \tfrac{1}{\eta}\right)[\beta_0 + \beta_l l_{jt} + \beta_k k_{jt}] - \left(-\tfrac{1}{\eta}\right)q_{mt} - h(\omega_{jt-1}) \end{aligned}$$

Let $\mathcal{F}_t$ be an observable subset of $\mathcal{F}_t^*$. Then equation (15) implies

$$E[\psi_{jt}(\boldsymbol{X}_{jt}, \boldsymbol{\beta}_a, h_a)|\mathcal{F}_t] = 0 \quad j = 1, \cdots, N \quad t = 1, \cdots, T \tag{16}$$

If the information set $\mathcal{F}_t$ is informative enough, such that $E[\psi_{jt}(\boldsymbol{X}_{jt}, \boldsymbol{\beta}, h)|\mathcal{F}_t] = 0$ for all j and for any $0 \leq \beta < 1$, then $(\boldsymbol{\beta}, h)^{'} = (\boldsymbol{\beta}_a, h_a)^{'}$. The true parameter values must satisfy the minimum distance relation

$$\boldsymbol{\alpha}_a = (\boldsymbol{\beta}_a, h_a)^{'} = arg \min_{\boldsymbol{\alpha}} E[m(\mathcal{F}_t, \boldsymbol{\alpha})^{'} m(\mathcal{F}_t, \boldsymbol{\alpha})]$$

where $m(\mathcal{F}_t, \boldsymbol{\alpha}) = E[\psi(\boldsymbol{X}_t, \boldsymbol{\alpha})|\mathcal{F}_t]$, $\psi(\boldsymbol{X}_t, \boldsymbol{\alpha}) = (\psi_1(\boldsymbol{X}_t, \boldsymbol{\alpha}), \cdots, \psi_N(\boldsymbol{X}_t, \boldsymbol{\alpha}))^{'}$ for any candidate values $\boldsymbol{\alpha} = (\boldsymbol{\beta}, h)^{'}$. The moment conditions are used to describe the SMD estimation of $\boldsymbol{\alpha}_a = (\boldsymbol{\beta}_a, h_a)^{'}$. The SMD procedure has three parts. First, we can estimate the function $h(\cdot)$, which has an infinite dimension of unknown parameters, by a sequence of finite-dimensional unknown parameters (sieves) denoted h_H. Approximation error decreases as the dimension H increases with sample size N. Second, the unknown conditional mean $m(\mathcal{F}_t, \boldsymbol{\alpha}) = E[\psi(\boldsymbol{X}_t, \boldsymbol{\alpha})|\mathcal{F}_t]$ is replaced with a consistent nonparametric estimator $\hat{m}(\mathcal{F}_t, \boldsymbol{\alpha})$ for any candidate parameter values $\boldsymbol{\alpha} = (\boldsymbol{\beta}, h)^{'}$. Finally, the function h_H is estimated jointly with the finite dimensional parameters $\boldsymbol{\beta}$ by minimizing a quadratic norm of estimated expectation functions,

$$\hat{\boldsymbol{\alpha}} = arg \min_{\boldsymbol{\beta}, h_H} \frac{1}{T} \sum_{t=1}^{T} \hat{m}(\mathcal{F}_t, \boldsymbol{\beta}, h_H)^{'} \hat{m}(\mathcal{F}_t, \boldsymbol{\beta}, h_H) \tag{17}$$

We approximate $h(\cdot)$ by a third order polynomial and substitute it in (16) as if it were the true model. Since the errors $\psi_t(\cdot)$ are orthogonal to the regressors $\mathcal{F}_t = (1, l_{t-1}, k_t, qm_{t-1}, x_{t-1})$, we use a third order power series of $\mathcal{F}_t$, denoted $\boldsymbol{P}$, as instruments. We estimate $m(\mathcal{F}, \boldsymbol{\alpha})$ as the predicted values from regressing the errors $\psi_t(\cdot)$ on the instruments. Using $\boldsymbol{P}$, we specify the weighting matrix as $\boldsymbol{A} = I_N \otimes (\boldsymbol{P}^{'}\boldsymbol{P})^{-1}$, making the estimation a GMM case. The weighting matrix $\boldsymbol{A}$ gives greater weight to moments that are highly correlated with the instruments. Using the specified GMM implementation, the parameter values $(\boldsymbol{\beta}, h_H)$ are jointly estimated.

Appendix C: Productivity decompositions

Though we cannot determine the exact contribution of large entrants, our data allow us to decompose aggregate productivity growth due to entrants, exits, and incumbents. Industry level productivity (Ω_t) can then be expressed as the weighted average productivity: $\Omega_t \equiv \sum_{j \in N} ms_{jt}\omega_{mt}$, where N is the number of stores, and $ms_{jt} = sales_{jt}/sales_t$.

The change in retail food productivity from year t to year t' can be written as

$$\begin{aligned} \Delta\Omega_{t,t'} = \ & \textstyle\sum_{j \in C_{t,t'}} ms_{jt}\Delta\omega_{jt,t'} + \sum_{j \in C_{t,t'}} \Delta ms_{jt,t'}(\omega_{jt} - \Omega_t) \\ & + \textstyle\sum_{j \in C_{t,t'}} \Delta ms_{jt,t'}\Delta\omega_{jt,t'} + \sum_{j \in E_{t,t'}} ms_{jt'}(\omega_{jt'} - \Omega_t) \\ & - \textstyle\sum_{j \in X_{mt,t'}} ms_{jt}(\omega_{jt} - \Omega_t) \end{aligned} \tag{18}$$

where Δ is the difference operator ($\Delta\Omega_{t,t'} = \Omega_{t'} - \Omega_t$); $C_{t,t'}$ is the set of continuing stores, i.e., operating in both t and t'; $E_{t,t'}$ is the set of entering stores, i.e., that operated in t' but not in t; and $X_{t,t'}$ is the set of exiting stores, i.e., that operated in t but not in t'. This decomposition, derived by Foster et al. (2001)(FHK), is a modified version of the decomposition by Baily et al. (1992).

The decomposition (18) thus consists of five terms. The first term (Within) is the increase in productivity when the continuing stores increase their productivity at initial sales. The second term (Between) is the increase in productivity when continuing stores with above-average productivity expand their share of sales relative to stores with below-average productivity. The third term (Cross) captures the increase in productivity when continuing stores increase their market shares, while the fourth and fifth terms (Entry and Exit) are productivity increases due to entry and exit, respectively.

The second productivity decomposition used is given by Griliches and Regev (1995) (GR) and modified by FHK to allow for entry and exit

$$\begin{aligned} \Delta\Omega_{t,t'} = \ & \textstyle\sum_{j \in C_{t,t'}} \overline{ms}_j\Delta\omega_{jt,t'} + \sum_{j \in C_{t,t'}} \Delta ms_{jt,t'}(\overline{\omega}_j - \overline{\Omega}) \\ & + \textstyle\sum_{j \in E_{t,t'}} ms_{jt'}(\omega_{jt'} - \overline{\Omega}) \\ & - \textstyle\sum_{j \in X_{t,t'}} ms_{jt}(\omega_{jt} - \overline{\Omega}) \end{aligned} \tag{19}$$

where bars over a variable indicate the average of the variable across t and t'. The within term in the GR decomposition consists of the growth rates of continuing stores' TFP weighted by the average of their shares across t and t'. Both decompositions compare aggregate productivity of entering and existing stores, either to the aggregate productivity of all stores (FHK) or to the unweighted average of aggregate productivity of all stores (GR).

Olley and Pakes (1996) (OP) proposes a static decomposition of aggregate productivity, in which the weighted productivity of continuing stores, Ω_t, has two components: (1) contribution of productivity improvements, $\overline{\Omega}_t$; and (2) market share reallocations for the continuing stores $cov(ms_{jt}, \omega_{jt}) \equiv \sum_j (ms_{jt} - \overline{ms}_t)(\omega_{jt} - \overline{\Omega}_t)$. The difference in productivity index,$\Delta\Omega_{t,t'}$, can be written as

$$\Delta\Omega_{t,t'} = \Delta\overline{\Omega}_{t,t'} + \Delta cov_{t,t'}. \tag{20}$$

The OP decomposition ignore the entry and exit. However, Melitz and Polanec (2009) (MP) suggest a dynamic OP decomposition where there is a positive contribution for entering and exiting stores only when the aggregate productivity of these stores is larger than that of continuing stores in corresponding periods. The aggregate productivity in periods t and t' can be decomposed as

$$\begin{aligned} \Omega_t &= ms_{C_t}\Omega_{C_t} + ms_{X_t}\Omega_{X_t} \\ \Omega_{t'} &= ms_{C_{t'}}\Omega_{C_{t'}} + ms_{E_{t'}}\Omega_{E_{t'}} \end{aligned} \tag{21}$$

where ms_{C_t}, $ms_{C_{t'}}$, $ms_{E_{t'}}$, and ms_{X_t} are the aggregate market shares of incumbents (in period t and t'), entrants and exits, respectively. The change in aggregate productivity can be written as

$$\Delta\Omega_{t,t'} = \Delta\overline{\Omega}_{C_{t,t'}} + \Delta cov_{C_{t,t'}} + ms_{E_{t'}}(\Omega_{E_{t'}} - \Omega_{C_{t'}}) + ms_{X_t}(\Omega_{C_t} - \Omega_{X_t}). \tag{22}$$

where the contribution of continuing firms is divided into within-firm productivity improvements ($\Delta\overline{\Omega}_{C_{t,t'}}$) and market share reallocations ($\Delta cov_{C_{t,t'}}$) as in OP. The contribution of entrants and exits contains two parts, unweighted average productivity (direct effect) and the covariance term (indirect effect). For entrants: $ms_{E_{t'}}(\overline{\Omega}_{E_{t'}} - \overline{\Omega}_{C_{t'}})$, and $ms_{E_{t'}}(cov(\Omega_{E_{t'}}) - cov(\Omega_{C_{t'}}))$. For exits: $ms_{X_t}(\overline{\Omega}_{X_t} - \overline{\Omega}_{C_t})$, and $ms_{X_t}(cov(\Omega_{C_t}) - cov(\Omega_{X_t}))$.