Finance Faculty Research and Publications   Finance, Department of

6-2014

Measuring the Impact of Agglomeration on Productivity: Evidence from Chilean Retailers

Sergio Garate
Marquette University, sergio.garate@marquette.edu

Anthony Pennington-Cross
Marquette University, anthony.pennington-cross@marquette.edu

Measuring the Impact of Agglomeration on Productivity: Evidence from Chilean Retailers

Sergio Garate

College of Business Administration, Department of Economics, Marquette University, Madison, Wisconsin

Anthony Pennington-Cross

College of Business Administration, Department of Finance, Marquette University, Milwaukee, Wisconsin

Abstract

This research extends the agglomeration literature to a country that has not been studied and a market sector that has received little attention. The majority of research that examines how density affects productivity has indirectly measured productivity through worker wages or property prices. The research uses individual supermarkets' store productivity, proxied by 10 years of annual sales per square foot. Studying supermarkets permits the examination of the effect consumers might have on productivity. Agglomerations (density) could

increase or decrease productivity depending on the relative extent of increased competition versus productivity gains, as consumers choose where to shop based on their interests in reducing shopping time (transport costs) and comparison shopping (product quality and pricing). Stores are described by who operates the store, the brand of the store and the size of the store. Results indicate that density has a differential impact depending on the store itself and the mix of stores nearby.

Keywords
agglomeration, consumer, retail

1. Introduction
Much of the empirical work related to measuring agglomeration economies (the productive benefits of density) measures productivity indirectly through observed wages or rents (Rosenthal and Strange, 2008; Combes *et al.*, 2008, 2011) or firm profitability (Jennen and Verwijmeren, 2010). In this paper, we examine the role of density or agglomerations on the productivity of individual supermarkets in Chile. Our proxy for productivity is the real peso value of sales per square metre for each store, measured annually for all supermarkets in Chile from 1997 through 2010. This is a significant departure from the prior literature, which has looked to input prices for evidence of different levels of productivity. Instead, we focus on the value of output normalised by the size of the store's physical infrastructure. The question asked in this paper is whether clustering of retail stores has any impact on productivity and, if so, what types of locations and mixes of competitors have the largest positive or negative spillovers. These results can help inform how to devise better development strategies that incorporate supermarkets. In fact, expenditures in the UK on food and non-alcoholic drinks are an increasingly important part of a family's budget and grew from 15.4 percent of total family expenditures to 16.5 percent over the 2005–10 time period (Office of National Statistics, 2011).

Agglomerations of supermarkets can be characterised as coming from at least two sources—production and consumption. The first source, production, includes the traditional microfoundations such as spillovers from worker interactions, historical decisions (where to place rail lines and ports, for example), the thickness of input markets and other factors. However, many of the factors that would promote productivity gains may be less important for a supermarket. For example, since most store sales positions require few skills there is likely to be little benefit from having a thick and skilled labour force. Consistent with this, Gabe and Abel (2011) find no evidence of benefits to low-skilled labourers from agglomeration.

The second source, consumption, includes non-traditional micro foundations focused on consumer demand for low cost and high information consumption. Co-location of similar supermarkets may help to reduce search costs and increase the number of potential customers (Picone *et al.*, 2009). The search may be focused on price or for particular goods or brands. Some households might find it advantageous to take a single shopping trip to pick up cheap toilet paper from the discount store and a fresh croissant or organic vegetables from the upscale grocer. This type of behaviour can extend the market area for all the stores in the location by creating a destination. While these stories of complementarity make sense, stores selling similar products will also act as substitutes and can just cannibalise a fixed demand in a certain location and lead to lower prices.

There is growing evidence that for shopping centres and large retailers, or big box retailers, agglomerations are more important in determining sales and entry than proxies of location productivity and even access to the transport network (Schuetz, 2013; Eppli and Shilling, 1996). Supermarket entry tends to decrease prices and employment, but the effect attenuates very quickly (Basker and Noel, 2009; Ellickson and Greico, 2013). In Chile, incumbent competitors tend to reduce prices before entry (Lira *et al.*, 2007).

As Chile's ability to distribute food has improved and consumer wealth has increased supermarkets have spread across the country. The number of storefronts increased from 430 to 1040 from 1997 to 2010. This period of rapid market growth and innovation provides a unique setting in which to study agglomeration effects. In addition, as a vestige of the Pinochet dictatorship (1973–90), firms can enter and exit in Chile very easily due to a weak regulatory environment and limited effective zoning (Zunino, 2006). This contrasts with the UK where entry of supermarkets is heavily restricted leading to 25 percent efficiency losses in England and Wales (Chesire *et al.*, 2011).

This paper contributes to the literature on agglomerations by focusing on the role of consumer-driven density in an important segment of consumption (food), using a unique dataset that covers the census of sales over a long time period, and examining all these issues in the context of Chile, which has been rapidly advancing during a long period of democratic and political stability in a weak regulatory environment.

The rest of the paper is organised as follows. Section 2 describes the model; section 3 discusses sources of bias and potential solutions; sections 4 and 5 present the data and empirical results; and section 6 presents our conclusions.

2. The Model

We analyse the elasticity of supermarket store-level sales productivity in Chile with respect to relative density, competitive conditions (concentration of some types of brands) and the variety of store types in the surrounding area. Direct measurement of store productivity would include output per square metre for each store. We observe the total value of the output (price times quantity) in each store over time and the size of the store. Our proxy for output is the real peso value of sales normalised by store size measured in square metres. Since output prices are likely to vary by store, it is important to control for input costs and profit margins which could affect output prices.

We use an annual dataset from 1997 to 2010 of each store that includes sales, the description of store sales area, brand and operator. Descriptions of the surrounding areas at each store include measures of the population and households, estimates of relative density or agglomeration and the extent of competition.

Under the assumption that agglomeration economies and a firm's specific production function can be expressed separately, we defined the production function as follows

$$y_{ij} = g(A_j)f(x_j) \qquad (1)$$

where y_{ij} is the real sales per square metre of sales area for store i in location j; $f(x_j)$ is the production function of store j; and x_j is a vector with store's capital, materials, labour and land all of them divided by square metres of sales area; and $g(A_j)$ is a shifter due to location-specific characteristics.

A more specific description of agglomeration driven productivity shifts was proposed by Rosenthal and Strange (2004)

$$A_j = \sum_{k \in K} q(y_j, y_k) a(d_{jk}^i, d_{jk}^t, d_{jk}^g) \qquad (2)$$

The quantity of activity in firm j and firm k is represented by $q(y_j, y_k)$. The industry distance between firm j activity and firm k activity is represented by d_{jk}^i. The temporal distance between firm j activity and

firm k activity is represented by d_{jk}^t. Agglomerations that took place in the past can have an impact on agglomerations and perhaps even productivity today. In Los Angeles, housing density and prices are higher today at locations where street car lines existed in the 1920s, even though they were removed by the 1940s and 1950s (Brooks and Lutz, 2012). The geographical distance between firm j and firm k is represented by d_{jk}^g.

Taking logs directly from equation (1) leads to

$$\ln(y_{ij}) = \gamma \ln(g(A_j)) + \delta \ln(f(x_j)) + \varepsilon_{ij} \quad (3)$$

where γ represents a vector of elasticities of sales productivity with respect to local characteristics and δ represents a vector of elasticities with respect to all inputs. We can rewrite this equation to control for agglomeration

$$\ln(y_{ij}(t)) = \boldsymbol{\alpha \Omega}_j(t) + \boldsymbol{\beta \eta}_j(t) + \boldsymbol{\theta \mu}_i(t) + \varepsilon_{ij}(t) \quad (4)$$

where $\ln(y_{ij})$ represents the sales productivity for store i and location j; $\boldsymbol{\Omega}_j$ describes the agglomerations for location j; $\boldsymbol{\eta}_j$ represents other location characteristics where the store is located; $\boldsymbol{\mu}_i$ is the specific production function for store i; ε_{ij} is a specific shock for firm i in location j; and t indexes time measures in years.

We measure store productivity, y_{ij}, as the sales per square metre of sales area in real Chilean pesos. Our measurement of agglomeration represents the fraction of supermarket activity in a location normalised by the fraction of the population in a location. More specifically agglomerations are defined as

$$\boldsymbol{\Omega}_j(t) = \frac{realsales_j(t) \Big/ realsales_{Chile}(t)}{population_j(t) \Big/ population_{Chile}(t)} \quad (5)$$

The numerator is the real pesos of sales in location j as a fraction of sales in the whole country. The location's share of sales is normalised by the location's share of the population. The primary definition of location j will be concentric rings around the store at 1000 metres and administratively defined municipalities.[1] The elasticity of productivity with respect to agglomeration is represented by α. Rosenthal and Strange (2003 and 2008) find that localisation economies attenuate rapidly in the first few miles and that human capital spillovers are largest close to college-educated workers. For retail sales, or supermarkets, the type of supermarket and the mix of types of supermarkets are also likely to affect the extent of spatial spillovers and productivity gains.

Store productivity also depends on the firm-specific production function, μ_j. The firm reflects a bundle of attributes that are observable, such as brand, format and operator and some others that are not easy to observe such as management in the store. The effect of these inputs can be controlled for with fixed effects, time-varying fixed effects, or information on these attributes.

Market characteristics of the store location, η_j, could create shifts in productivity. Urban areas receive a 33 per cent wage premium relative to the non-urban areas (Glaeser and Maré, 2001). The effect of location on store productivity can be controlled for using fixed effects, time-varying fixed effects, or information that is specific to the location. Location reflects the 'natural' attributes or other attributes that define the character and shopping

potential of the location. The error term ε_{ij} is a shock specific to store i and in location j in year t that are independent and identically distributed.

3. Potential Sources of Bias

Estimates of agglomeration economies must deal with at least three major challenges—measurement error, endogenous explanatory variables and non-random selection. Fortunately, our measure of productivity and agglomerations is derived from a census of storefront sales and is unlikely to be measured with much (if any) error. Agglomerations are endogenous—for example, firms may be attracted to locations with higher productivity, thus increasing agglomerations. Sorting can occur; that is, stores can non-randomly enter and exit a location, depending on firm characteristics. For example, a firm may make different strategic entry and exit choices depending on the type or format of the store involved.

3.1 Selection and Endogenous Agglomeration

The main purpose of this study is to identify the extent to which agglomerations, Ω_j, impact store productivity. Ordinary least squares estimates of α provide an unbiased estimate only when the covariance, as defined here, is zero

$$\text{Cov}(\boldsymbol{\Omega}_j, \boldsymbol{\mu}_i + \boldsymbol{\eta}_j + \varepsilon_{ij}) = 0 = \text{Cov}(\boldsymbol{\Omega}_j, \boldsymbol{\mu}_i) + \text{Cov}(\boldsymbol{\Omega}_j, \boldsymbol{\eta}_j) + \text{Cov}(\boldsymbol{\Omega}_j, \varepsilon_{ij}) \tag{6}$$

Agglomerations are correlated with the store type: $Cov(\boldsymbol{\Omega}_j, \boldsymbol{\mu}_i)$

Store location is not likely to be randomly selected. Real estate developers have information on competitors and seek locations with higher productivity to enter. Therefore, higher store productivity may attract more businesses thus increasing agglomerations.

Agglomerations are correlated with the location fixed effects: $\text{Cov}(\boldsymbol{\Omega}_p, \boldsymbol{\mu}_p)$

A 'natural' advantage or location attribute that increases the sale per square foot for the store and also increases the agglomeration of sales will bias the estimate of α upwards. This could include factors such as unobserved bus stops or changes in transport infrastructure.

Agglomerations are correlated with store sales shocks: $\text{Cov}(\boldsymbol{\Omega}_p, \varepsilon_{ip})$

For example, if the frequency of a store experiencing positive demand shocks depends in part on agglomerations then the covariance will be positive. This may occur in dense and diversified shopping districts where the next big fad or popular toy is more likely to be available. These shocks then increase foot traffic and sales for all stores and hence the extent of agglomerations.

Common approaches used to reduce covariance issues

A common approach to these issues is to use various forms of fixed effects (Glaeser and Maré, 2001). These will control for any unobserved location and store characteristics that do not vary over time and are related to the selection of the location. In our case, we can control directly for the type of store by identifying the format, the brand, the operator and store size. The location defines the potential customer base and the availability of a labour force. Identification is aided by the entry and exit of stores and time variations in sales. Over the 1997 through 2010 time period, every municipality experienced at least some entry and in total over 700 stores opened (entered). Since entry and exit is unlikely to be random, the estimation must identify what these stores look like and the impact of store characteristics. Location fixed effects imply that α is identified by changes over time in the agglomeration and cross-sectional variation in agglomerations. Location fixed effects will confound identification of α when location attributes affect sales and agglomerations at the same time. Examples include

the building of a new road or major construction on local access streets or highway off-ramps. One approach we will use to limit this problem is to include individual store fixed effects.

Agglomerations and sales productivity may be jointly determined. Instrumental variables are one approach to help identify the direct effect of agglomeration on sales productivity. Good instruments must be capable of explaining agglomerations but not productivity. One potential type of instrument is population or location information from a long time ago (long lags). These variables may be reasonable if they are related to the persistence of spatial density but are not related to productivity contemporaneously. Following the spirit of other studies (Ciccone and Hall, 1996; Combes *et al.*, 2008, 2011; Mion and Nattichioni, 2009; Rosenthal and Strange, 2008; and Redding and Venables, 2004), we use long lags of population. We use the population census from 1885. Identification is aided because these measures come before the Pinochet dictatorship (1973–90) which introduced a very open economy mixed with brutal suppression of the population's political and economic points of view, transforming Chile socially and economically. Strong economic growth did not take hold until the transition to democracy began in the early 1990s. In short, our instruments cover a time period of substantially different economic regimes and the country has undergone huge changes and growth since 1885.

In an effort to identify highly productive places in the past, we use the distance to the central municipality in each region according to 1907 census. The reference point used in each of these municipalities was the central 'plaza' if the municipality was located in the central valley of Chile or the customs office if the municipality was located on the coast of Chile. Most of the cities in the central valley were located in highly fertile land, which would have driven population density and agricultural business activity in the 1800s. Now, the soil fertility has no relationship with economic activity because the land is fully developed and is used in non-agricultural activities. One example is Santiago's central plaza, founded in 1541 and located on the fertile lands of the central valley, 500 metres from the Mapocho River. Today, downtown Santiago is dominated by financial services.

3.2 Two-step Market Area (Municipality) Approach

The empirical approaches discussed so far measure the impact of location characteristics on individual store productivity. The use of concentric rings around each store creates a market area or location that is unique for each store. This awkwardly mixes together in the fixed effects attributes of the location and the store. An alternative approach is to use market areas or shopping destinations/districts as the definition of location. This separates unobserved location and store attributes. This requires estimating time-varying market area specific productivity measures and examining whether the location's productivity is related to agglomerations and other factors. For identification, this requires a two-step approach

$$\text{Firststep: } y_{ijt} = \boldsymbol{\theta}\boldsymbol{\mu}_i + \boldsymbol{\beta}\boldsymbol{\eta}_{jt} + \varepsilon_{ijt}$$

$$\text{Secondstep: } \boldsymbol{\beta}_{jt} = \alpha\boldsymbol{\Omega}_{jt} + \psi\boldsymbol{\lambda}_{jt} + \tau\boldsymbol{T}_t + \boldsymbol{\phi}\boldsymbol{X}_t + \varepsilon_{2jt} \tag{7}$$

The individual store sales productivity equation (y_{ijt}) is estimated in the first step. Explanatory variables are store fixed effects (μ_i) and time-varying market area fixed effects ($\boldsymbol{\eta}_{jt}$). In the second step, the estimated time-varying market area productivity is explained by agglomerations ($\boldsymbol{\Omega}_{jt}$), a vector of other location variables including municipality fixed effects ($\boldsymbol{\lambda}_{jt}$), other variables ($\boldsymbol{X}_t$) and time dummies ($\boldsymbol{T}_t$). This specification is similar in spirit to Combes *et al.* (2008, 2011) and Mion and Nattichioni (2009). One key feature of this approach is that permanent store characteristics are controlled for.

4. Data Description

In 2012, Chile had an estimated population of 17,402,630 (INE, 2012)—about the same order of magnitude as the population of the US state of Florida in 2010. In 2011, Chilean GDP was US$234 billion—about the same order of magnitude as the US state of Indiana's gross product. Chile's GDP grew on average at a rate of 4.01 per cent over the 2004–11 time period.[2] Chile is divided into regions, provinces and finally into 346 communes or municipalities. The northern regions are dominated by mining industries, while the rest of the country has manufacturing, transport, communication personal services and agriculture. In the metropolitan region, where Santiago is located (the country's capital), the main sectors of the economy are financial services, commerce, restaurants and hotels. Together these account for approximately 50 percent of the Metropolitan Region's output.[3]

In 2002, supermarket sales in Chile represented 60 per cent of the retail food market, which is well below the ratio found in developed countries. The penetration of supermarkets is 75 percent in Denmark, 70 percent in Norway, 80 percent in Sweden, 88 percent in the UK and 90 percent in the US (Traill, 2006). According to the National Institute of Statistics, between July 2010 and July 2011, supermarket sales in Chile grew 9.3 percent in nominal terms and 8.4 percent in real terms. The number of supermarkets also increased for the same period from 1067 to 1169. This trend has been a characteristic of this market for some time, with a yearly average increase in the number of stores of 5.2 percent and in nominal sales growth of 7.4 percent for the period 2003–10 (INE, 2011).

The sales information comes from a database provided by one of the competitors in the supermarket industry and is produced by the National Institute of Statistics. The dataset begins in 1997 with 434 stores and ends in 2011 with 1171 stores. Each store is geocoded using the store address. For publicly traded companies, sales area was collected from press releases. For the remaining 581 stores on-site observation was used (Smu, 2012; Cencosud, 2012; Walmart, 2012). To control for any error in the observation of sales area, we added a dummy variable (*Arq*) to differentiate the first group from the latter. For each store and year, the brand, format and operator are identified.

As part of the analysis, we need to control for the market concentration of competitors as well as the type of store. By the end of 2010, store operators include

1. Walmart Chile, with a sales market share of 36.5 per cent and 269 buildings.
2. Cencosud, with a sales market share of 27.8 percent and 158 buildings.
3. Supermercados Unimarc, with a sales market share of 22.4 percent and 274 buildings.
4. Tottus, with a sales market share of 5.9 percent and 30 buildings.
5. Others with a total sales market share of 7.2 percent and 199 buildings.

Store brand can differ from the operator and a single operator can run multiple brands. There are also a variety of formats that each of the stores can use. Store formats or store fronts can be categorised as follows (Smu, 2012; Cencosud, 2012; Walmart, 2012)

- —Convenience store: on average 1700 stock keeping units (sku) and 250 square metres of sales area.
- —Traditional supermarket: 16,000 sku and 1550 square metres of sales area.
- —Discount/wholesale store: limited assortment, 4500 sku and 2000 square metres of sales area.
- —Hypermarket: 50,000 sku and 7400 square metres of sales area.

All of these stores have limited general merchandise (other than food) that represents no more than 5 per cent of sales with the exception of hypermarkets that have close to 20 per cent of sales. In 2010, the formats include supermarkets with 601 stores, discount stores with 244 stores, convenience with 64 stores and hypermarkets with 113 stores. Traditional supermarkets look very much like supermarkets in the US (similar to Safeway). Discount stores look like warehouses selling goods with open boxes and bins for consumers to sort through (similar to Costco, but with more of an emphasis on processed food and less customer service). Convenience stores are much smaller and cater to local consumption (similar to 7-Eleven). Hypermarkets are very large stores that sell a variety of goods but still focus of food (fresh and processed). This supermarket segment is dominated by Walmart affiliates (similar to big box retailers). In 2010, traditional supermarkets had the largest market share of sales (55 percent) and hypermarkets were not too far behind (41 percent). Discount stores and convenience stores had, respectively, much lower market shares of 3.6 percent and less than 1 percent. Over the 1997 through 2010 time period, all of the formats were growing rapidly but hypermarkets have been gaining substantial market share from traditional supermarkets with a growth rate over 400 percent for hypermarkets and a growth rate over 65 percent for traditional supermarkets.

We constructed the Herfindahl–Hirschman index (Hhi) to control for the market power of the operator. We also construct a divergence index (div) to control for the variety of store formats (the industry distance between firm j activity and firm k activity).

$$
\begin{aligned}
\text{hhi}_l &= \sum_{j}^{n} \text{marketshare}_{jl}^2 \\
\text{div}_l &= \sum_{k}^{n} \text{marketshare}_{kl}^2
\end{aligned}
\tag{8}
$$

Market share in the Hhi index represents the market share of operator j in location l. Market share for the div index represents the market share of store format k in location l. For the Hhi index, a value of one indicates monopoly, only one operator, and zero indicates an infinite number of firms with no significant market share. Similarly, for the div index one indicates that all the stores in the area are of one format.

Tables 1 and 2 describe the variables. The variables are divided into store-level information and municipality-level information used in the two-step approach. Tables 3 and 4 provide the descriptive statistics for each of the variables. There is substantial variation in the productivity of stores ranging from just over 13,000 pesos per square metre to 1.8 million pesos per square metre. Consistent with the variety of store formats, the size of stores ranges from 100 to over 13,000 square metres. Walmart (*Op4*) is the largest operator, with 15 percent market share (based on store-years). Brand, when based on store years, is likewise very diverse, with the largest brand (Santa Isabel, *Br9*) holding a 15 percent market share. The largest format is Traditional supermarkets with a 77 percent market share. Within a format, brand operations can become very concentrated. For example, for traditional supermarkets in 2010, operators Cencosud (*Op3*) and Supermercados Unimarc (*Op2*) have market shares of approximately 29 percent each; and, for convenience stores, Resto (*Op6*, part of 'other') has a market share above 75 percent.

Table 1. Description of variables for store level variables

Variable	Description variable	Unit

Basic information		
Ω	Supermarket agglomerations within a 1000 metre concentric ring around the store	Index
Hhi	Operator market power within a 1000 metre concentric ring around the store	Index
Div	Format variety or divergence within a 1000 metre concentric ring around the store	Index
Nominal Sales	Nominal sales of the store for each year	Chilean pesos (CLP)
Sales/Mt2	Productivity measured as real sales per square metre; CLP of deflated by the national consumer price index	Real CLP/square metre
Mt2	Sales area of the store	Square metre
Arq	0 if company reported square metres and 1 if observed square metres	
Store operator		
Op1	1 if it is operated by Unimarc, 0 if not	
Op2	1 if it is operated by the Supermercados Unimarc (SMU), 0 if not	
Op3	1 if it is operated by the Cencosud, 0 if not	
Op4	1 if it is operated by Walmart, 0 if not	
Op5	1 if it is operated by Southern Cross, 0 if not	
Op6	1 if it is operated by Others, 0 if not	
Store brand		
Br1	1 if Brand is Express de Lider, 0 if not	
Br2	1 if Brand is Hiper de Lider, 0 if not	
Br3	1 if Brand is Unimarc, 0 if not	
Br4	1 if Brand is Mayorista 10, 0 if not	
Br5	1 if Brand is Montserrat, 0 if not	
Br6	1 if Brand is Jumbo, 0 if not	
Br7	1 if Brand is Santa Isbael, 0 if not	
Br8	1 if any other brand, 0 if not	
Store format		
Fconv	1 if format of store is Convenience, 0 if not	
Fsuper	1 if format of store is Traditional Supermarket, 0 if not	
Fhyper	1 if format of store is Hypermarket, 0 if not	
Fdisc	1 if format of store is Discount, 0 if not	
Instrumental variables		
Census population		
c1885	Population in 1885 census within a 1000 metre concentric ring from the store	People
Distance		
Dist	Distance to main ports and downtowns from the store in 1907	Kilometres

Table 2. Description of variables for municipality level data

Variable	Description variable	Unit
Municipality level data		
Ω_{muni}	Municipality retail activity density relative to Chile's density	Index

HHimuni	Municipality Herfindahl–Hirschman Index (*Hhi*) in terms of real sales	Index
Divmuni	Municipality divergence index in terms of real sales	Index
(Sales/Mt2)muni	Productivity measured as real sales per square metre; CLP of deflated by the national consumer price index	Real CLP/square metre
Mt2muni	Municipality average store size	Square metres
Instrumental variables		
Census population		
c1885muni	Municipality population in 1885 census	People
Distance		
Dist1muni	Average distance to main ports and downtowns of each census block within a municipality in 1907.	Kilometres
Dist2muni	Minimum distance to main ports and downtowns of each census block within a municipality in 1907	Kilometres

Table 3. Descriptive statistics: store-level variables

Variable	Mean	S.D.	Minimum	Maximum
Basic information				
Ω	0.16	0.44	0.00	15.46
Hhi	0.67	0.26	0.18	1.00
Div	0.88	0.18	0.34	1.00
Sales/Mt2	329,899.70	201,325.10	13,108.76	1,875,984.00
Mt2	1806.29	2206.75	100.00	13,243.00
Arq	0.48	0.50	0.00	1.00
Operator information				
Op1	0.04	0.21	0.00	1.00
Op2	0.08	0.27	0.00	1.00
Op3	0.12	0.32	0.00	1.00
Op4	0.15	0.35	0.00	1.00
Op5	0.03	0.16	0.00	1.00
Op6	0.58	0.49	0.00	1.00
Store brand				
Br1	0.05	0.21	0.00	1.00
Br2	0.07	0.26	0.00	1.00
Br3	0.08	0.27	0.00	1.00
Br4	0.02	0.15	0.00	1.00
Br5	0.03	0.18	0.00	1.00
Br6	0.02	0.14	0.00	1.00
Br7	0.15	0.36	0.00	1.00
Br8	0.58	0.49	0.00	1.00
Format information				
Fconv	0.05	0.21	0.00	1.00
Fsuper	0.77	0.42	0.00	1.00
Fhyper	0.11	0.31	0.00	1.00
Fdisc	0.07	0.26	0.00	1.00
Instrumental variables				
c1885	16,355.93	32,160.91	1.00	177,271.00

Dist	21.70	27.89	0.04	184.91

Note: The number of observations for store-level data is 8640.

Table 4. Descriptive statistics: municipality variables

Variable name	Mean	S.D.	Minimum	Maximum
Municipality-level data				
$\Omega_{muni,Chile}$	0.96	0.87	0.00	7.93
Hhimuni	0.66	0.27	0.15	1.00
Divmuni	0.86	0.20	0.34	1.00
(Sales/Mt2)muni	298,560.10	136,756.60	18,128.70	1,456,50.00
Mt2muni	1391.19	1182.23	150.00	8278.00
Instrumental variables				
c1885muni	10,497.55	17,882.38	1.00	17,7271.00
Dist1muni	30.93	29.77	0.69	183.97
Dist2muni	28.31	29.67	0.00	182.05

Note: The number of observations for municipality-level data is 2121.

Measures of format (Div) and operator (Hhi) concentration using a 1000 metre concentric ring show that there is more concentration in format than in operators. The instrumental variables show that there was a substantial variation in population in 1885 and distance to economic centres (ports and downtowns) in 1907. The municipality variables echo these findings.

5. Results
Since there are repeated observations for each location, we relax the assumption of independence of observations and use the Huber and White sandwich estimator of variance (Petersen, 2009). In addition, we take the log of all continuous variables so that the coefficients can be interpreted as elasticities.

5.1 Basic Store Results
Table 5 presents ordinal least squares (OLS) results for the store-level approach. All the specifications include store-level fixed effects. The first column, specification I, includes the basic controls and year fixed effects. The results are largely statistically significant and the regression has good explanatory power (R^2 is over 80 percent), indicating that agglomerations (Ω) are on average positively associated with store-level productivity. Larger stores (*Mt2*) are associated with lower productivity, perhaps indicating the inability of big stores to capture any productivity spillovers. Store size also will be a strong proxy for the format of the store. More market power of the operator (*Hhi*) or the format (*div*) is associated with more productive stores. Specification II includes a format dummy variable with traditional supermarkets as the reference group. The format dummies are largely insignificant indicating that store size (Mt2) and store-level dummies are likely to provide very good proxies for format. Specification III includes brand and operator dummy variables with other being the excluded category for both. There is little evidence that the operator makes much difference in store productivity. Brand does seem to matter with Express de Lider and Santa Isabel being the most productive brands. However, the explanatory power of the results is not greatly improved and these additional controls have no meaningful impact on the role of agglomerations.

Table 5. Store ordinary least squares results (dependent variable: ln(*Sales/Mt2*))

		I			II			III	
Category	*Variable*	*Coefficient*	*S.E.*		*Coefficient*	*S.E.*		*Coefficient*	*S.E.*
Basic	ln(Ω)	0.32***	0.03		0.32***	0.03		0.31***	0.03

	ln(Hhi)	0.16*	0.10	0.17*	0.09	0.17*	0.09
	ln(Div)	0.08*	0.05	0.08	0.05	0.08*	0.05
	arq	-0.13*	0.08	-0.13*	0.07	-0.08	0.07
	ln(Mt2)	-0.81***	0.19	-0.84***	0.16	-0.87*	0.13
Operator	Op1					0.20	0.06
	Op2					0.05	0.04
	Op3					0.06	0.04
	Op4					-0.22	0.13
	Op5					-0.04	0.05
Brand	Br1					0.42***	0.14
	Br2					0.08	0.14
	Br3					-0.11**	0.05
	Br4					-0.09	0.07
	Br5					-4.05***	0.15
	Br6					0.78*	0.33
	Br7					-0.02	0.06
format	Fconv			0.10	0.09		
	Fhyper			0.36*	0.20		
	Fdisc			-0.39	0.27		
Year	1998	0.02*	0.01	0.02*	0.01	0.02*	0.01
	1999	-0.03*	0.02	-0.04*	0.02	-0.04*	0.02
	2000	-0.05**	0.02	-0.05***	0.02	-0.05***	0.02
	2001	-0.06**	0.02	-0.06***	0.02	-0.06***	0.02
	2002	-0.07***	0.03	-0.07***	0.03	-0.07***	0.03
	2003	-0.06*	0.03	-0.06*	0.03	-0.06*	0.03
	2004	-0.04	0.03	-0.04	0.03	-0.05	0.04
	2005	-0.03	0.04	-0.03	0.04	-0.04	0.04
	2006	-0.02	0.04	-0.02	0.04	-0.03	0.04
	2007	-0.01	0.04	0.00	0.04	-0.01	0.04
	2008	0.03	0.04	0.03	0.04	0.02	0.05
	2009	0.05	0.05	0.05	0.05	0.05	0.05
	2010	0.08*	0.05	0.08*	0.05	0.07	0.05
	Constant	19.19***	1.19	19.36***	0.99	19.35	0.86
	Store fixed effects	y		y		y	
Observations		8640		8640		8640	
Stores		863		863		863	
R^2		0.8341		0.8356		0.8377	

Notes: * $p<0.1$; ** $p<0.05$; *** $p<0.01$. Excluded categories are: operated by others, brand other, format traditional supermarket, and year 1997. Concentric rings are 1000 square metres. T-statistics are clustered by store and robust.

Table 6 indicates that the OLS results were biased as the estimated coefficient for the instrumented agglomeration drops from 0.31 to 0.05 and becomes statistically insignificant. Again, store-level fixed effects are included, but agglomerations are instrumented using population within 1000 metres of the store in 1885 and the distance to main ports and downtowns in 1907. The bottom panel provides statistics on how the instruments performed as a tool for identification. The instruments provided some additional explanatory power of agglomerations beyond the other control variables with an R^2 of 0.12 and a statistically significant F-statistic

of 3.90. The null hypothesis of weak and underidentification is rejected using Kleibergen–Paap tests. The Hansen J-statistic does not reject the null that the instruments are valid. In sum, the instruments seem to be reasonable and provide modest explanatory power. However, this should be expected given the 100-year time interval and the time-invariant nature of the instruments.

Table 6. Store instrumental variables (IV) results (dependent variable: ln($Sales/Mt2$))

Category	Variable	Coefficient	SE
Shopping	ln(Ω)	0.05	0.07
	ln(Hhi)	0.19**	0.09
	ln(Div)	-0.14*	0.07
	arq	-0.22***	0.07
	ln(Mt2)	-0.02	0.10
Operator	$Op1$	0.31***	0.06
	$Op2$	0.10***	0.04
	$Op3$	0.08*	0.04
	$Op4$	-0.40**	0.16
	$Op5$	-0.14	0.14
Brand	$Br1$	0.61***	0.19
	$Br2$	0.78***	0.24
	$Br3$	-0.07	0.06
	$Br4$	-0.03	0.07
	$Br5$	0.26	0.19
	$Br6$	0.68***	0.18
	$Br7$	0.12	0.09
	Constant	14.04***	0.47
	Year dummies	y	
	Store fixed effects	y	
Observations		8640	
Stores		863	
R^2		0.7095	
First-stage statistics			
1st-stage F-statistic on excluded instruments			3.90***
Partial R^2 of excluded instruments			0.12
Weak identification test			
Kleibergen–Paap Wald rk F statistic (Chi-squared)			33.44**
Underidenitfication test			
Kleibergen–Paap rk LM statistic (Chi-squared)			80.04***
Overidentification test of all instruments			
Hansen J-statistic			1.41

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Excluded categories are: operated by others, brand other, format wholesaler, and year 1997. T-statistics and identification tests are clustered by municipality and robust. Excluded instruments are the log of population in 1885 within a 1000 metre concentric ring around the store and a 5000 metre concentric ring, and the log of the distance to the nearest main ports and downtown in 1907.

The results also indicate that the operator is a more important determinate of store productivity than indicated in the OLS results. In particular, Unimarc operated stores are the most productive and Walmart operated stores are the least. In addition, store size is no longer significant and the sign switches on format market power. The coefficient for operator market power is still positive.

5.2 Basic Municipality Results

These results follow the specification of the two-step approach discussed in section 3. In the first step, the specification explains store productivity using dummy variables for each store and time variant dummy variables for each municipality. The time-variant municipality coefficients provide an estimate of municipality productivity over time which controls for the effect of the stores located in the municipality. This municipality measure of productivity is then explained in a second step by municipality agglomerations (Ω_{muni}), municipality market power of the operator and format (HHI_{muni} and Div_{muni}), year dummies, average store size in the municipality ($Mt2_{muni}$), the average age of stores in the municipality (Age_{muni}), year dummies and municipality fixed effects. To help account for the precision of the left-hand variable, each observation is weighted by the inverse of the first-stage coefficient's standard errors.

The first column of Table 7 reports the OLS results and the second column reports the instrumental variables (IV) results (Ω_{muni}). Both approaches provide good explanatory power with R^2s over 0.84. Format market power or concentration is associated with lower productivity. Operator market power has no effect and neither does the age of the building. Similar to the store results, the IV results show no relationship between agglomerations and productivity, while the OLS results find a positive and significant elasticity. The excluded instruments partial R^2 of 0.27 and a statistically significant F-statistic indicate that the instrumental variables seem to function adequately. Again, the null hypothesis of weak and underidentification is rejected and the Hansen J-statistic does not reject the null that the instruments are valid.

Table 7. Municipality results (dependent variable: predicted ln(*Sales/Mt2*)*muni*)

Variable	OLS		IV	S.E.
	Coefficient	S.E.	Coefficient	
ln(Ω_{muni})	0.22***	0.05	0.00	0.11
ln(Hhi_{muni})	0.01	0.09	-0.08	0.10
ln(Div_{muni})	0.20**	0.10	0.22**	0.09
ln($Mt2_{muni}$)	-0.07	0.12	0.01	0.12
ln(age_{muni})	0.00	0.00	0.00	0.00
Constant	-0.44	0.68	-1.13	0.78
Year dummies	y		y	
Municipality fixed effects	y		y	
Observations	2112		2112	
Municipalities	177		177	
R^2	0.8683		0.8484	
First-stage statistics				
1st-stage F-statistic on excluded instruments				7.71***
Partial R^2 of excluded instruments				0.27
Weak identification test				
Kleibergen–Paap Wald rk F statistic (Chi-squared)				20.14**
Underidenitfication test				
Kleibergen–Paap rk LM statistic (Chi-squared)				45.24***

Overidentification test of all instruments				
Hansen J-statistic				5.74

5.3 Differential Impacts of Agglomerations

Table 8 asks the question of whether agglomerations have differential impacts on productivity by store format. Using the IV approach, the first and second groups of columns report results using the store-level and municipality approaches. The sample is limited in each row to only stores of one format and then the specification is rerun on that sub-sample.[4] Half of the coefficients are statistically significant and never across both estimation approaches. However, there is some consistency in the sign of the coefficients. Agglomerations tend to reduce productivity for discount/warehouse stores and traditional supermarkets. Both of these store formats are large, stand-alone or anchor types of stores which are viewed by property managers and developers as stores that bring in customer traffic but do not necessarily benefit from customer traffic generated by other businesses. By contrast, convenience stores rely on being in the right location when customers are already shopping. Therefore, as expected, the coefficient on agglomerations is positive and significant for the municipality specification for convenience stores. This may reflect the importance of 'shopping districts' or destinations when measuring agglomerations for secondary retail (non-anchor).

Table 8. IV results for overall agglomerations by store format (dependent variable: ln(*Sales*/*Mt2*))

Store format	Store ln(Ω)			Municipality ln(Ω_{muni})	
	Coefficient	*S.E.*		*Coefficient*	*S.E.*
Convenience	0.05	0.07		0.36**	0.15
Hypermarket	0.19**	0.09		-0.48	0.30
Discount/warehouse	-0.14*	0.07		-0.30	0.48
Traditional supermarket	-0.22***	0.07		-0.18	0.30

Notes: * p <0.1; ** p <0.05; *** p <0.01. Each reported coefficient represents the impact of agglomerations for all store formats on a sub-sample that is limited to the indicated store format. T-statistics are clustered by municipality and robust. The specification from Table 6 is used for the store results and Table 7 for the municipality results.

Table 9 further examines the differential impact across property formats by asking whether the agglomeration of specific formats has differential impacts on certain types of stores. Applying the sub-sampling approach used in Table 8, each row represents a separate regression for the sub-sample using one type of supermarket agglomeration at a time.[5] Less than half of the coefficient estimates are statistically significant. The strongest and most consistent results indicate that hypermarket productivity is lower when there is a higher density of discount/warehouse within 1000 metres or within a municipality. This result is symmetrical— discount/warehouse stores also have lower productivity when there is a density of hypermarkets. In short, these types of format should not co-locate and are likely strong substitutes for each other.

Table 9. IV Results for format agglomerations by store format (dependent variable: ln(*Sales*/*Mt2*))

Store Format	Store ln(Ω_{format})		Municipality ln($\Omega_{muni\ \&\ format}$)	
	Coefficient	S.E.	Coefficient	S.E.
Hypermarket Ω				

Traditional supermarket	0.71	0.64	0.31	0.24
Convenience	-0.90	0.78	0.13	0.25
Hypermarket	0.28	0.21	-0.16***	0.02
Discount/warehouse	-0.11***	0.04	-0.12***	0.08
Traditional Supermarket Ω				
Traditional supermarket	-0.05	0.11	-0.18	0.31
Convenience	-0.01	0.08	0.45***	0.10
Hypermarket	-0.02	0.05	0.10	0.07
Discount/warehouse	0.07**	0.04	0.18	0.34
Convenience Ω				
Traditional supermarket	-0.21	0.21	-0.93	1.16
Convenience	0.13	0.08	-0.10	0.16
Hypermarket	2.15*	1.28	-0.28	0.33
Discount/warehouse	0.30**	0.14	-0.16	0.13
Discount/warehouse Ω				
Traditional supermarket	0.46	0.39	-1.06	1.25
Convenience	-0.25	0.62	0.24**	0.01
Hypermarket	-0.21***	0.54	-0.11***	0.04
Discount/warehouse	0.11	0.11	-0.40*	0.23

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Each reported coefficient represents the impact of agglomerations for a specific store format on a sub-sample that is limited to the indicated store format. T-statistics are clustered by municipality and robust. The specification from Table 6 is used for the store results. The specification from Table 7 is used for the municipality results.

6. Conclusion

In this paper, we analyse the effect of agglomeration on supermarket productivity. The data cover all supermarkets in Chile from 1997 through 2010. Store characteristics such as size, operator, format, brand, sales and location are observed. The store formats range from small convenience stores to huge box retailers called hypermarkets (huge supermarkets are often operated by Walmart) and discount/warehouse stores that sell packaged food and goods out of bins and stacks with almost no customer service. In between are traditional supermarkets.

The results indicate that, on average, the density or agglomeration of other supermarkets around an individual store has no measureable impact on productivity. One interpretation of these results is that there are very few productivity spillovers for supermarkets. An alternative interpretation is that firms locate in locations until the benefits of the density equal the costs of density.

The benefits for supermarkets are likely to be derived from consumption not production. This consumption-driven agglomeration also allows for more comparison shopping in terms of price and quality. At some point, density creates congestion thus increasing transport and opportunity costs of shopping. It also reduces prices. The overall agglomeration results are consistent with firms entering until all positive effects are dissipated through congestion and pricing effects. This interpretation of the results is aided by the weak regulatory environment of Chile, which allows entry and exit with little interference.

However, when the effect of agglomerations is allowed to vary by store format, differential impacts do appear. For example, the evidence indicates that convenience stores have higher productivity when there is an agglomeration of all other supermarkets in a shopping district of destination. This impact is largest when the agglomeration is from traditional supermarkets and discount/warehouses store formats. This finding is consistent with the idea that some types of store are destinations and drive their own traffic (for example,

discount/warehouses), while other types of store benefit by being near the primary destinations. In contrast, some types of store have no complementary interaction. For example, the discount/warehouse store format and the hypermarket format both reduce the productivity of each other almost symmetrically. There is also consistent evidence that stores of the same format tend to cannibalise sales from each other.

These spatial interactions of store types and types of agglomeration suggest several recommendations or findings that should be useful for urban policy. First, if entry is left unregulated, supermarkets (as broadly defined here) will continue to enter and create agglomerated supermarket shopping until all productivity gains are dissipated. Second, despite the fact that big box stores tend to locate near one another (Schuetz, 2013), the co-location of large supermarkets (hypermarkets and discount markets) should be discouraged due to high rates of sales cannibalisation. These stores are best dispersed among the population they serve. Third, small supermarkets should be encouraged to locate near larger and traditional supermarkets to take advantage of the increased traffic flow.

In summary, the evidence is consistent with the idea that the market will naturally find the limits of the benefits of supermarket density. However, within different formats there are still differential impacts of density and hence co-location. Sometimes the productivity spillovers are positive and sometimes they are negative.

Funding

Notes
1.Results at 500 metres to 5000 metres were tested and are available from the authors on request.
2.See: Central Bank Chile website,
    at http://si3.bcentral.cl/Siete/secure/search/ViewBasket_NEW.aspx (accessed 27 June 2012).
3.See note 2.
4.For the municipality data, municipalities are identified as heavy in traditional supermarkets if all the supermarkets are traditional, as heavy in convenience, hypermarket or discount/warehouse if the municipality has any of these types of format.
5.It is necessary to run each format's agglomeration measure separately because we do not have the ability to meet the exclusion requirements for four different endogenous measures of agglomerations simultaneously using long lagged demographic and distance data.

References

Basker, E., Noel, M. (2009) The evolving food chain: competitive effects of Wal-Mart's entry into the supermarket industry, Journal of Economics and Management Strategy, 18(4), pp. 977–1009.
Brooks, L., Lutz, B. (2012) Vestiges of transit: path dependence and the modern city. Paper presented at the American Real Estate and Urban Economics Mid-year Conference, May.
Cencosud (2012) Investor relation (www.cencosud.cl/eng/index.htm; accessed March 2012).
Chesire, P., Hilber, C., Kaplanis, I. (2011) Evaluating the effects of planning policies on the retail sector: or do town centre first policies deliver the goods? Discussion Paper No. 66, Spatial Economic Research Center (www.spatialeconomics.ac.uk/textonly/SERC/publications/download/sercdp0066.pdf; accessed 4 June 2013).
Ciccone, A., Hall, R. E. (1996) Productivity and the density of economic activity, American Economic Review, 86(1), pp. 54–70.
Combes, P.-P., Duranton, G., Gobillon, L. (2008) Spatial wage disparities: sorting matters!, Journal of Urban Economics, 63(2), pp. 723–742.

Combes, P.-P., Duranton, G., Gobillon, L. (2011) The identification of agglomeration economies, Journal of Economic Geography, 11(2), pp. 253–266.

Ellickson, P., Greico, P. (2013) Wal-Mart and the geography of grocery retailing, Journal of Urban Economics, 75, pp. 1–14.

Eppli, M., Shilling, J. (1996) How critical is a good location to a regional shopping center?, The Journal of Real Estate Research, 12(3), pp. 459–468.

Gabe, T., Abel, J. R. (2011) Agglomeration of knowledge, Urban Studies, 48(7), pp. 1353–1371.

Glaeser, E. L., Maré, D. C. (2001) Cities and skills, Journal of Labor Economics, 19(2), pp. 316–342.

INE (Instituto Nacional de Estadística Chile) (2011) Índice de Ventas de Supermercados 2011 (www.ine.cl/canales/chile_estadistico/estadisticas_economicas/supermercados/series_estadisticas/series_estadisticas.php).

INE (2012) Demografia y Vitales: Instituto Nacional de Estadísticas (www.ine.cl/canales/chile_estadistico/demografia_y_vitales/demo_y_vita.php).

Jennen, M., Verwijmeren, P. (2010) Agglomeration effects and financial performance, Urban Studies, 47(12), pp. 2683–2703.

Lira, L., Rivero, R., Vergara, R. (2007) Entry and prices: evidence from the supermarket sector, Review of Industrial Organization, 31(4), pp. 237–260.

Mion, G., Naticchioni, P. (2009) The spatial sorting and matching of skills and firms, Canadian Journal of Economics, 42(1), pp. 28–55.

ONS (Office of National Statistics) (2011) Family spending 2011 edition, a report on living costs and food survey 2010 (www.ers.usda.gov/data-products/food-expenditures.aspx; accessed 4 June 2013).

Petersen, M. A. (2009) Estimating standard errors in finance panel data sets: comparing approaches, Review of Financial Studies, 22(1), pp. 435–480.

Picon, G., Ridley, D., Zandbergen, P. (2009) Distance decrease with differentiation: strategic clustering by retailers, International Journal of Industrial Organization, 27(3), pp. 463–473.

Redding, S., Venables, A. J. (2004) Economic geography and international equality, Journal of International Economics, 62(1), pp. 53–82.

Rosenthal, S. S., Strange, W. C. (2003) Geography, industrial organization, and agglomeration, The Review of Economics and Statistics, 85(2), pp. 377–393.

Rosenthal, S. S., Strange, W. C. (2004) Evidence on the nature and sources of agglomeration economies, in: Henderson, J. V., Thisse, J.-F. (Eds) Handbook of Regional and Urban Economics, pp. 2119–2171. Amsterdam: Elsevier.

Rosenthal, S. S., Strange, W. C. (2008) The attenuation of human capital spillovers, Journal of Urban Economics, 64(2), pp. 373–389.

Schuetz, J. (2013) Why are Wal-Mart and Target next-door neighbors? Paper presented to the Homer Hoyt Institute, May.

Smu (Unimarc Supermarkets) (2012) Investor relation (www.smu.cl/; accessed March 2012).

Traill, W. B. (2006) The rapid rise of supermarkets?, Development Policy Review, 24(2), pp. 163–174.

Walmart (2012) Investor relation (www.dys.cl/wps/wcm/connect/dys/DyS/inversionistas/informacion_financiera/MemoriasW/; accessed March 2012).

Zunino, H. (2006) Power relations in urban decision-making: neo-liberalism, 'techno-politicians' and authoritarian redevelopment in Santiago, Chile, Urban Studies, 43(10), pp. 1825–1846.