

# Hotelling meets Holmes:

The importance of returns to product differentiation and distribution economies  
for the firm's optimal location choice

Anett Erdmann\*

Universidad Carlos III de Madrid

This Version: March 2013

## Abstract

Inspired by the empirical work of Holmes (2011), which suggests the economic importance of distribution costs in the firm's optimal location decision, this paper introduces endogenous distribution costs in the model of Hotelling (1929). The proposed model shows an interesting trade-off between demand and cost considerations when a firm plays a *mixed-location strategy*. Given the location of local distribution centers and agents' displacement cost parameters, it is shown that, under certain conditions, the optimal location of the firms are in the interior of the Hotelling line rather than at the edges of the line. The supply cost effect which drives this result diminishes with the distance of the distribution center from the market so that the scale of the distribution area becomes also determinant for an optimal location strategy. The theoretical results are complemented with an empirical analysis for distribution intensive grocery retailers using location data for the two main conventional supermarket chains in the U.S. The data suggest that the firms consider distribution costs when differentiating from the competitor.

**Keywords:** Product differentiation, distribution costs, grocery retailers.

**Contact details:** Anett Erdmann, Universidad Carlos III de Madrid, Department of Economics, Calle Madrid 126, 28903, Getafe (Madrid), Spain; Email: aerdmann@eco.uc3m.es

---

\*This paper is the first project within my Ph.D. dissertation at University Carlos III de Madrid. I thank my advisors, Makoto Watanabe (TI) and Jesús Carro (UC3M), for their support and guidance during this as well as ongoing projects. I also want to thank Ulrich Wagner (UC3M) and Ricardo Flores-Fillol (URV) for their comments and suggestions. Any remaining errors are my own.

# 1 Introduction

For distribution intensive industries with strong cost focus and high turnover rate of merchandise, business concepts suggest that the optimization of logistic costs plays a crucial role to be competitive (Andersen and Poulsen, 2006). However, most of the economic models either do not account at all for distribution costs or include them (implicitly) as exogenous fixed costs. Such a setting is in general unproblematic for market entry models but may be problematic in the context of optimal geographic differentiation between competing firms. A certain location decision considering demand and competition effect may be optimal for given fixed costs but once we consider supply costs, as part of the fixed costs and depending on the actual location of the firm and its distribution center, it might be profitable to locate closer to the distribution facility to decrease supply costs although this may imply less differentiation to competitors. Inspired by the empirical work of Holmes (2011) which suggests the economic importance of distribution costs in the optimal location decision of a firm (Wal-Mart), this paper introduces endogenous distribution costs in the duopoly model of Hotelling (1929).

To the best of my knowledge, this tension between demand and supply strategy of location choice has neither been analyzed in a theoretical model of product differentiation nor is there any empirical analysis for oligopoly industries.

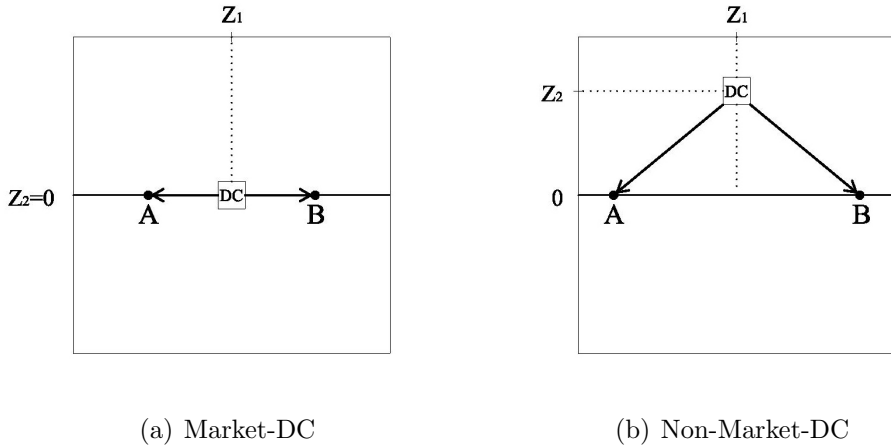
The theoretical work-horses of spatial location choice are Hotelling's linear city model (1929) and the subsequent work by d'Aspremont et al.(1979) which allow to analyze product differentiation in a simple framework. There is a large literature modifying or extending the Hotelling framework, for a review see Anderson et al.(1992) or Tirole (1998). More recent examples are Meagher et al.(2008) analyzing the equilibrium existence under different consumer distributions or Hamoudi and Moral (2005) considering linear-quadratic transportation costs for consumers. But while the demand side has been extensively analyzed, to the costs of the firms, in particular product specific fixed costs, has not been paid much attention. The theoretical literature is complemented by empirical and computational papers. For given supermarket locations, Matsa (2011) shows that the distance to the distribution centers has a negative effect on the store's product availability. Considering endogenous location choice, in the context of market entry decisions, Mazzeo (2002) and Seim (2006) provide empirical evidence for the market power effect of product differentiation within a market. But except in the fixed entry costs, also here, supply side aspects have been a long time unconsidered. The first empirical analysis incorporating supply aspects in an endogenous location choice model is Holmes (2011). He uses a computational analysis to show in a dynamic market entry model that the optimal location strategy of Walmart is based on a trade-off between proximity of stores to distribution centers and own store demand cannibalization. His work has inspired other researchers to control for supply distances in empirical market entry models (e.g. Ellickson, 2010).

This paper proposes a price-location game in which firms use mixed location strategies considering cost-efficiency and horizontal competition simultaneously. On the demand side consumers incur in travel costs to buy at a certain store. On the supply side, each firm's store is day-to-day stocked up by an (own) exogenous distribution center, which can be located in or outside the linear market, and firms have to bear the supply costs. Consumers face quadratic travel costs while firms' displacement costs are modeled as a linear-quadratic function of the supply distance to allow for a more flexible shape, since contrary to consumers, suppliers are allowed to 'travel' to firms from outside the market.

Solving for the optimal location choice, shows an interesting trade-off between returns to product differentiation and distribution economies. It is shown that under certain conditions, depending on the location of the distribution centers and agents' displacement

egy. On the demand side consumers incur in travel costs to buy at a certain store. On the supply side, each firm's store is day-to-day stocked up by an exogenous distribution center, which can be located in or outside the linear market, and firms have to bear the supply costs.<sup>3</sup> The location of the DC in space is characterized as  $(z_1^a, z_2^a)$  and  $(z_1^b, z_2^b)$  respectively. We use a reference coordinate system where the linear market builds the horizontal-axis and the left end of the market is defined as the origin of the coordinate system.<sup>4</sup> Hence, the shortest distance from the DC to the market can directly be indicated as  $|z_2^j|$  and the orthogonal projection of a DC onto the market is just  $(z_1^j, 0)$ , with  $j = \{a, b\}$ . Figure 1 illustrates exemplarily two possible situations where both firms are supplied by a common DC. While in Figure (a) the firms are supplied by a DC which is situated in the linear market ( $z_2 = 0$ ), Figure (b) illustrates a situation where firms are stocked up by a DC located outside the market. In the following I refer to this two cases as *Market-DC* and *Non-Market-DC* respectively.

Figure 1: The Linear City with Distribution Costs



Considering the consumer side, consumer  $i$  who lives at  $x_i$  faces quadratic travel costs  $TC_i(a) = t \cdot (a - x_i)^2$  if he buys from A and  $TC_i(b) = t \cdot (r - b - x_i)^2$  if he buys from B, where  $t$  is the travel cost parameter and  $a$  and  $r - b$  the respective firm locations. Firms' displacement costs are specified in a similar way. Distribution costs are modeled as linear-quadratic functions of the supply distance which can be reduced to a quadratic function for the simple case where the DC is located inside the market. We choose this cost specification to allow for a more flexible shape, since contrary to consumers, suppliers are allowed to 'travel' to firms from outside the market. The distribution distance can be simply expressed as the hypotenuse of a right-angled triangle between the DC and

<sup>3</sup>The exogeneity assumption of the DCs is easy to justify whenever the DCs belong to a third party or a firm leases already existing DCs of another chain (Recent example: Target entering the Canadian market). If the DCs are firm own, the DC-location may be considered as an endogenous decision of the firm. In this paper we abstract from this special case, focusing on firms that use ex-ante established facilities or third party service providers.

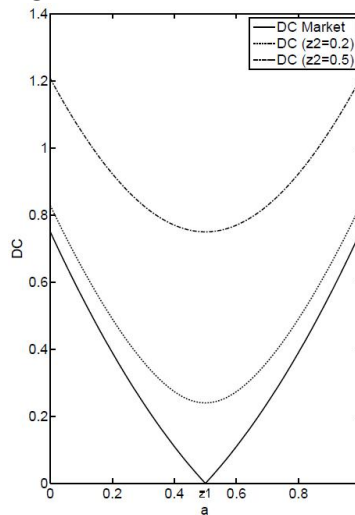
<sup>4</sup>A similar framework of firms choosing their locations on a line while the environment is allowed to be two-dimensional is used by Thomadsen(2006). He places two heterogeneous fast food stores on a line and let them choose their optimal locations in terms of the distance from the center while consumers are distributed over a two-dimensional space.

the store location. Hence, given the location of firm A's DC or exogenous supplier, the distribution costs for store A are given by

$$\begin{aligned} DC_a(a; z^a) &= d_1 \cdot (\text{Supply Distance})^2 + d_2 \cdot (\text{Supply Distance}) \\ &= d_1 \cdot [(a - z_1^a)^2 + (z_2^a)^2] + d_2 \cdot \sqrt{(z_1^a - a)^2 + (z_2^a)^2} \end{aligned}$$

where  $d_1$  and  $d_2$  are distribution cost parameters capturing the linear-quadratic shape of the supply cost function. To reduce the analysis to a non-negative, increasing and convex supply cost function, I assume  $d_1 \geq 0, d_2 \geq 0$ .<sup>5</sup> The specification includes the case of quadratic costs ( $d_2 = 0$ ) which is illustrated in Figure 2. It is immediately clear that the distribution costs increase with the distance of the distribution center from the market. But notice that the distribution cost effect of moving one unit closer to the projected distribution facility ( $z_1$ ) increases the closer the supplier is to the market (smaller  $z_2$ ).

Figure 2: Distribution costs



$$d_1 > 0, d_2 = 0$$

With this in mind, the firms' strategic decisions take place in two stages. First, firm A and B decide simultaneously their store locations  $a$  and  $r - b$  respectively. The feasibility constraint of the location choice, which is indicated in terms of the distance from the market edges, implies that  $a, b \in [0, r]$ . Additionally, I assume that  $a \leq b$ . Once the two grocery firms are established they compete in prices. The firms set prices  $p_a$  and  $p_b$  depending on the degree of differentiation and hence the strength of competition in the market. Based on each firm's location and the prices offered, the utility maximizing consumers face a discrete choice problem at which store to buy.

---

<sup>5</sup>The linear-quadratic cost specification in the Hotelling model is not new. Hamoudi and Moral (2005) for example use a linear-quadratic cost specification for consumers' travel costs to allow for concave transportation costs. We use a similar specification for the supply costs but impose the restriction of a convex cost structure. Instead of assuming  $d_1 \geq 0$ , the assumption could be relaxed allowing for a non-monotonic shape of the cost function. In this case, in order to guarantee a non-negative cost function, I may extend the cost specification to a general second degree polynomial  $DC_a(a; z^a) = d_0 + d_2 \cdot SDistance + d_1 \cdot (SDistance)^2$  with at most one root. The additional term  $d_0 \geq 0$  could be interpreted as fixed operation costs.

## 2.2 Equilibrium locations

Given the environment presented in the previous section, the game is solved recursively. Compared with the standard linear city model, the pricing stage doesn't change and hence I abstain from a detailed discussion of this stage.<sup>6</sup> The indifferent consumer is given by  $\tilde{x} = \frac{p_b - p_a}{2t(r-b-a)} + \frac{r-b+a}{2}$ , the store-demand is  $D_a = \tilde{x}$  and  $D_b = r - \tilde{x}$  respectively so that the resulting optimal prices given any two store locations  $(a, r-b)$  are  $p_a^*(a, b) = c + t \cdot (r-a-b)r + 1/3 \cdot t \cdot (r-a-b)(a-b)$  and  $p_b^*(a, b) = c + t \cdot (r-a-b)r - 1/3 \cdot t \cdot (r-a-b)(a-b)$ . Given the optimal pricing decision and the exogenous locations of the DCs, firm A chooses its optimal location solving the following problem:

$$\left\{ \begin{array}{ll} \text{Max}_a [p_a^*(a, b) - c] D_a(p_a^*(a, b)) - DC_a(a; z_1^a, z_2^a) \\ \text{s.t.} & a \in [0, r] \end{array} \right\}$$

Note that the firm's location choice enters not only in the demand, but as well in the cost function. Solving for  $a$ , under the first order condition of the pricing stage, yields the following optimality condition for the firm's location choice:

$$FOC_a : \underbrace{(p_a - c) \left[ \frac{\partial D_a}{\partial a} + \frac{\partial D_a}{\partial p_b} \frac{\partial p_b}{\partial a} \right]}_{MRPD} \leq \underbrace{\frac{\partial DC_a}{\partial a}}_{MRDE} \quad (1)$$

The inequality condition (1) clearly indicates the trade-off between marginal returns on product differentiation (MRPD), which reflects the competition effect, and marginal returns in form of distribution economies (MRDE). It captures the dual nature of location choice and its effect on firm's profit. In other words, assuming  $a < z_1^a$ , if firm A moves marginally away from the extreme towards firm B, competition increases and revenues decrease but at the same time the firm moves closer to the distribution center, such that the firm saves on logistic costs.<sup>7</sup> If the savings on supply costs are bigger than the loss of revenues, it is optimal for the firm to move towards the competing firm at the cost of stronger price competition.

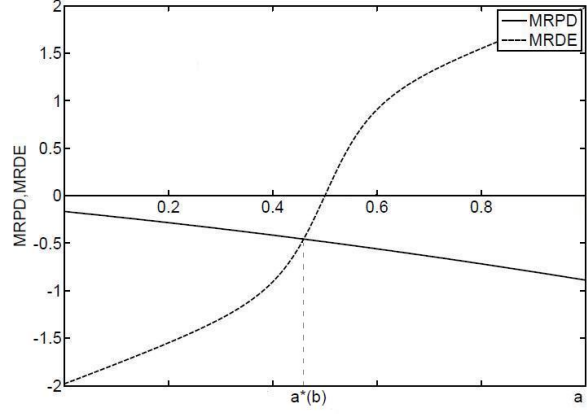
Figure 3 illustrates this trade-off. The profit is maximized where MRPD equals MRDE. The optimal location choice depends on the one hand on firm B's location choice and consumers' travel cost parameter which shift the MRPD and on the other hand on the firm's distribution cost parameters as well as the location of the DC which alter the degree of convexity of the MRDE.<sup>8</sup> However, the inequality in the best response condition (1) indicates that there may be situations where firm B's location and the set of displacement parameters is such that firm A chooses a corner solution locating at the market edge  $a = 0$ . Proposition 1 provides conditions which guarantee a best location response inside the market where MRPD equals MRDE with a view to the symmetric location choice.

<sup>6</sup>This is due to the setting analog Holmes (2011), defining the distribution costs as fixed costs independent of the sales volume. (In an extension we may additionally allow the unit variable costs to be an increasing function of the distribution distance, so that the DC location determines directly the optimal pricing decision.)

<sup>7</sup>Note that  $a > z_1^a$  can never be an optimal location for the firm, since a marginal decrease in  $a$  implies a reduction in distribution costs as well as an increase in market power.

<sup>8</sup>Exemplarily in Figure 3 I set  $b = 0, 3$  and the set of displacement parameters ( $t = d_1 = d_2 = 1$ ) with a DC at  $(z_1^a, z_2^a) = (0, 5; 0, 1)$ .

Figure 3: Best Location Response



**Proposition 1.** *The firm's best location is an interior solution on the Hotelling Line if the consumers' travel cost parameter is small enough (relative to the distribution cost parameter).*

In other words, a threshold value  $t_{crit}$  determines when the supply side consideration becomes relevant for the firm's location choice (see Appendix). Let's focus in the following on the interesting case where  $t < t_{crit}$ .

The effect of the location of the DC on the firm's optimal location response can be broken down into a *local-effect* and a *scale-effect*.

$$\frac{\partial DC_a}{\partial a} = - \left( \underbrace{2d_1 \cdot (z_1^a - a) + d_2}_{local-effect} \cdot \underbrace{\frac{(z_1^a - a)}{\sqrt{(z_1^a - a)^2 + (z_2^a)^2}}}_{scale-effect} \right)$$

The local-effect is the MRDE if the DC were located at the orthogonal projection of the DC on the linear market  $(z_1^a, 0)$ . This hypothetical location is used to identify the impact of the size of the distribution area which I denote as the scale-effect. The scale-effect is the part of the MRDE which is determined by the distance of the DC to the market. The latter is especially relevant if we think of logistic centers being located in industrial areas outside the town.<sup>9</sup>

**Proposition 2.** *The Supply-cost-effect on the optimal location response diminishes with the distance of the DC from the Market.*

The more far away the DC is with respect to the market (shopping area), the smaller is the scale-effect and hence the absolute value of the MRDE. This implies that the supply effect on the firm's optimal location choice is most relevant if the distribution center is not too far from the market. The importance of this result lies in the dependence of the firm's strategic location choice on the scale of its distribution area. Notice that the MRDE is zero when A settles down at the projected location of the DC, i.e. at  $a = z_1^a$ . However, this would only be an optimal location if we consider only the supply side ignoring the demand side incentive of product differentiation to create market power. In the following, I solve for the equilibrium considering both, supply and demand side implications of location

<sup>9</sup>An alternative argument are firms operating in several markets and being supplied by only one DC (not captured in the model).

choice. Considering firm A and B simultaneously yields a system of best responses. To solve for the optimal location choice, as mentioned before, I distinguish between the situation where the DC(s) are located inside the market and a more general situation allowing the DC(s) to be located outside the market.

### 2.2.1 Distribution center(s) inside the market

It is helpful to first consider the case in which the DC(s) are located somewhere on the Hotelling line. We refer to this as *Market-DCs*, since the distribution centers are located inside the market such that  $z_2^a = z_2^b = 0$ . From the optimal location condition as outlined in equation (1) we get a best-response-system  $BR_a(b), BR_b(a)$  which yields the following polynomial system:

$$\begin{aligned} a^2 \left(-\frac{1}{6}t\right) + a \left(-\frac{1}{2}t - \frac{1}{18}tr - 2d_1\right) + b \left(-\frac{1}{6}t + \frac{1}{18}tr\right) + b^2 \left(\frac{1}{18}t\right) + ab \left(\frac{1}{9}\right) + (2d_1z_1^a + d_2 - \frac{1}{6}tr) &= 0 \\ b^2 \left(-\frac{1}{6}t\right) + b \left(-\frac{1}{2}t - \frac{1}{18}tr - 2d_1\right) + a \left(-\frac{1}{6}t + \frac{1}{18}tr\right) + a^2 \left(\frac{1}{18}t\right) + ab \left(\frac{1}{9}\right) + & \\ (2d_1(r - z_1^b) + d_2 - \frac{1}{6}tr) &= 0 \end{aligned}$$

We can see from the polynomial structure that under DC-symmetry, that is if  $z_1^a = r - z_1^b$ , which includes the case of a co-located or joint DC at  $z_1 = \frac{1}{r}$ , we will have a symmetric location solution.<sup>10</sup> In the following I focus on the symmetric location equilibrium.<sup>11</sup> Solving for the optimal location choice yields the following symmetric and unique Nash-equilibrium:

$$a^*(z, t, d) = b^*(z, t, d) = \begin{cases} 0 & \text{if } t \geq t_{crit}(d, z), \\ (12d_1z_1^a + 6d_2 - tr) / (4t + 12d_1) & \text{if } t < t_{crit}(d, z), \end{cases} \quad (2)$$

The optimal location choice is characterized by the location of the DCs, captured in the vector  $z$ , as well as the displacement cost parameters  $t$  and  $d = (d_1, d_2)$ . Analog Proposition 1, we can express the threshold of an interior solution as a critical value of consumers' transportation costs  $t_{crit}$  (or as function of the relative importance of transportation and distribution costs captured in  $\gamma = t/d_1$  which requires  $\gamma < \gamma_{crit} = \frac{12z_1^a + 6}{r}$



demand side perspective). It is easy to show that if the DC is sufficiently far from the market edges and  $t < t_{crit}$ , the optimal location choice of the firm exists and is a unique interior point of the Hotelling line which solves the trade-off between MRPD and MRDE. A special case of an interior solution is the situation where  $a = z_1^a$  minimizes distribution costs.

The result is consistent with standard models. If  $t = 0$ , such that there is no incentive to differentiate geographically, firms settle down at the location of their distribution centers to minimize costs ( $z_1^a = z_1^b = r/2$  implies Bertrand's equilibrium). The other extreme are pretty high transportation costs for consumers. If the travel cost parameter  $t$  exceeds the critical threshold, which happens if  $t$  is much higher than the distribution cost parameter  $d_1$ , the demand strategy becomes dominant and firms choose maximal differentiation. Finally, if  $d_1 = d_2 = 0$ , the optimal location is again the one of the d'Aspremont case. Hence, the presence of firms' distribution costs can decrease the degree of product differentiation which enhances price competition.

### 2.2.2 Generalization of the DC location(s)

Let us now consider the more general case where firms are supplied from distribution centers which are allowed to be located outside the market. We now distinguish between *Market-DCs* and *Non-Market-DCs*, where the latter refers to DCs which are located off the Hotelling line. Given the location of the exogenous DC(s) and the displacement parameters of consumers and firms, Firm A's implicit best response on B's location choice becomes the following:

$$a^2 \left(-\frac{1}{6}t\right) + a \left(-\frac{1}{2}t - \frac{1}{18}tr\right) + b \left(-\frac{1}{6}t + \frac{1}{18}tr\right) + b^2 \left(\frac{1}{18}t\right) + ab \left(\frac{1}{9}\right) + \left(-\frac{1}{6}tr\right) \leq -2d_1(z_1^a - a) - d_2 \frac{(z_1^a - a)}{\sqrt{(z_1^a - a)^2 + (z_2^a)^2}}$$

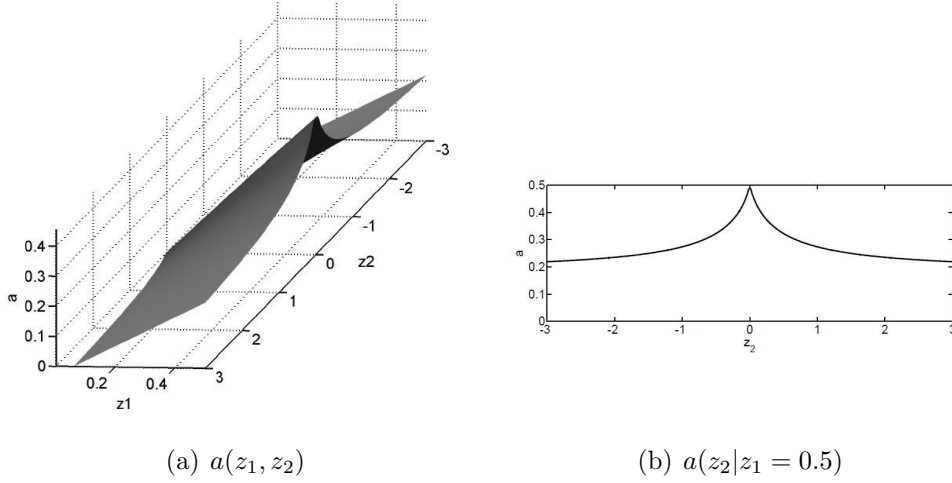
Firm B faces an analog trade-off. Note that whenever  $z_2 \neq 0$  (Non-Market-DC), the MRDE are not any more linear in  $a$ . This is due to the diminishing supply-cost-effect as stated in Theorem 2. This implies that the chosen cost specification of the distribution costs is intuitive but comes at a cost, that the model is no longer analytically solvable. However, focusing on the symmetric case of location choice, i.e.  $z_1^a = r - z_1^b$  and  $z_2^a = z_2^b = z_2 \in \mathbb{R}$ , the solution can be plotted for any set of displacement parameters  $(t, d)$ .

Figure 4a illustrates the dependence of the location choice  $a$  on the location of its DC at  $(z_1^a, z_2^a)$ . Since I focus on the symmetric case, the graph depicts only the market side for firm A (mirrored for firm B). It is easy to verify that, analog the previous section, the further away the DC projection  $(z_1^a, 0)$  is from the market edge, the larger is also  $a$ . But, considering the transverse section of the graph, depicted in Figure 4b, note that the effect diminishes in  $|z_2|$ , that is with the distance from the market.

Analog the case of Market-DC(s), the set of parameters for which an interior solution exists for the general case is defined as follows:

$$\Theta^* = \left\{ (z, d, t, r) : z_1^a \in [0, r]; d_1, d_2 \in \mathbb{R}^+; t, r \in \mathbb{R}^+; \frac{d_2 z_2^a}{\sqrt{(z_1^a)^2 + (z_2^a)^2}} + 2d_1 z_1^a \leq \frac{1}{6}tr \right\}.$$

Figure 4: Market-DC and Non-Market-DC



Although I can't solve for the general case analytically, the graphical illustration on the one hand confirms the results from the previous section and on the other hand exposes the impact of the size of the distribution area in the location consideration. That is, once firms consider distribution costs in their location decision, it may not be optimal anymore to employ maximal differentiation but the distribution economies which drive this result diminish when the distance from the DC to the market becomes large. In other words, the closer the DC is to the market, the stronger is the trade-off between returns to product differentiation and returns in form of distribution economies.

### 2.3 Welfare implications

To push the analysis further, I consider the welfare implications when accounting for distribution costs in the optimal location strategy. Maximizing social welfare in the Hotelling framework is equivalent to minimizing costs. However in the presented model there are two types of costs. While consumers' transportation costs are minimized at  $a^{TC} = b^{TC} = \frac{r}{4}$ , distribution costs are minimized at the projected DC location, that is at  $a^{DC} = b^{DC} = z_1^a = r - z_1^b$ . It is easy to deduce that in the interval  $[\frac{r}{4}, z_1^a]$  the social planner will face a trade-off between increasing total transportation costs and decreasing distribution costs, or inversely if  $z_1^a < \frac{r}{4}$ . In the following I solve for the social optimum for the case where the DC is located in the market to make it comparable to the closed form solution provided in section 2.2.1. The planner faces the following problem:

$$\text{Min} \{T(a, b) + D(a, b)\}_{(a, b)}$$

where  $T(a, b)$  are the total transportation costs paid by the consumers, that is  $T(a, b) = \int_0^{\tilde{x}} t(a-x)^2 f(x) dx + \int_{\tilde{x}}^r t(r-b-x)^2 f(x) dx$ , and  $D(a, b)$  are the total distribution costs paid by the firms so that  $D(a, b | z_2 = 0) = d_1 \cdot [(z_1^a - a)^2 + (r - b - z_1^b)^2] + d_2 \cdot [(z_1^a - a) + (r - b - z_1^b)]$ . Solving for the optimal locations yields the following first order condition for  $a$ :

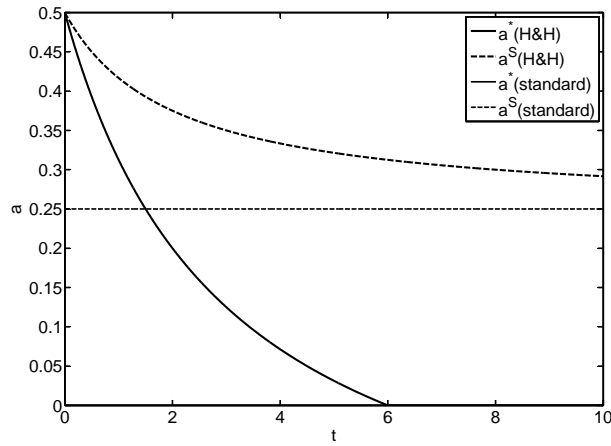
$$\frac{\partial T(a, b)}{\partial a} + \frac{\partial D(a, b)}{\partial a} = t \cdot \left( a^2 - \left( \frac{r-b-a}{2} \right)^2 \right) - d_1 \cdot (2z_1^a - 2a) - d_2$$

and analog for  $b$  with  $r - z_1^b = z_1^a$ . Solving for the social optimum, under the coexistence of positive travel costs and positive supply costs, the system yields after rearrangement the following social optimal locations:

$$a_{H\&H}^{social}(z, t, d > 0) = \frac{0.25tr^2 + 2z_1d_1 + d_2}{2d_1 + tr} \quad \text{and} \quad b_{H\&H}^{social} = r - a_{H\&H}^{social} \quad (3)$$

Note that while in the standard Hotelling framework without distribution costs the social optimum is independent of the consumers' transportation cost parameter ( $t > 0$ ), in my model which internalizes distribution cost effects, the social optimum depends on the displacement parameters of consumers and firms.

Figure 5: Optimal market location and social optimum (for firm A).



$$\theta_S = (z_1 = 0.5, z_2 = 0, d_1 = 1, d_2 = 0).$$

Moreover note that  $\lim_{(d_1, d_2) \rightarrow 0} a_{H\&H}^{social} = \frac{r}{4}$  which is consistent with the standard Hotelling setting. On the other hand  $\lim_{t \rightarrow 0} a_{H\&H}^{social} = z_1 + \frac{d_2}{d_1}$  which minimizes distribution costs. Setting  $d_2 = 0$ , which imposes quadratic distribution costs but is no problem whenever the DC is located inside the market as in the present case, the optimal location is just next to the DC. Finally, comparing the social optimum with the market outcome I find that the market forces still lead to excessive differentiation, that is a gap between market outcome and social optimum, but less than in the standard model if the supply cost parameter is sufficiently high relative to the consumers' transportation cost parameter. We briefly illustrate the excessive differentiation ( $\Delta D$ ) as a function of the relative importance of distribution costs and transportation costs defining  $\gamma \equiv t/d_1$  with  $t > 0$ ,  $d_1 > 0$ . I choose this representation since it reflects the relative importance of the competition effect which is the source of the inefficiency. Since this section considers the case of Market-DCs, without loss of generality I set  $d_2 = 0$  and  $r = 1$  such that the differentiation-gap is given by

$$\Delta D(\gamma, z_1^a) = a_{H\&H}^S - a_{H\&H}^* = \frac{2\gamma^2 + (5 - 4z_1)\gamma}{4\gamma^2 + 20\gamma + 24} \in \begin{cases} [0, \frac{1}{4}] & \text{if } \gamma \leq \bar{\gamma}(z_1^a), \\ [\frac{1}{4}, \frac{1}{2}] & \text{if } \gamma > \bar{\gamma}(z_1^a), \end{cases} \quad (4)$$

Consequently, for  $\gamma \leq \bar{\gamma}(z_1^a)$ , where  $\bar{\gamma}(z_1^a) = 2z_1^a + \sqrt{(2z_1^a)^2 + 6}$ , the gap between social optimal differentiation and the market outcome is smaller than in the standard model. Figure 5 illustrated this result and provides at the same time a comparison

of the market outcome (solid lines) and social optimum (dashed line) for the 'Hotelling meets Holmes' (H&H)-model and the standard Hotelling model as a function of consumers' transportation costs. Given any distribution cost setting  $\theta_S$ , the graph shows that the discrepancy between optimal market location and social optimum increases in  $t$  for interior solutions and decreases if  $t \geq t_{crit}$ , where the firm chooses a corner solution. In the limit ( $t \rightarrow \infty$ ) the supply effect is dominated by the competition effect and we are back to the standard model.

### 3 An application to the location of supermarkets

In this section I aim to verify the impact of distribution costs on firms geographic differentiation empirically for a particular example of a distribution intensive industry. I consider the leading conventional supermarket chains in the US, namely The Kroger Co. followed by Safeway Inc., both market-listed and operating predominantly 'neighborhood grocery stores'. With focus on the trade-off between differentiation from competitors and distribution economies, I have chosen competitors which are on a par with each other and abstract from the competitive pressure of mass merchandisers like Walmart on traditional supermarkets (see for example Jia, 2008, or Matsa, 2011).<sup>13</sup> I also abstract from a possible trade-off between differentiation and agglomeration which is considered by Datta and Sudhir (2011). I present first the data and the measure of differentiation and subsequently use a multivariate regression analysis to verify if the presented model offers a valid explanation for the firm behaviour revealed through the observed location choice.

#### 3.1 Data

I use data on supermarket locations in the U.S. for Kroger and Safeway. All the store locations have been identified from POI datasets.<sup>14</sup> The advantage of this type of data source is that locations are already geocodified which avoids matching problems in a manual geocodification (which would be necessary to measure efficiently the geographic differentiation between a huge number of stores). Additional information from the respective firm's website allow to identify the store format which is operated under a certain banner and the location of the regional distribution centers. Moreover stakeholder information, especially the 'Fact Book' and 'Annual Report', allow to verify the consistency of the POI data which turns out to be highly accurate. The differences in the number of stores indicated by the POI dataset and the official financial publications are 3 stores for Kroger and 24 stores for Safeway. Both small deviations with respect to the total number of stores of the chains. The difference is assigned to the time difference in the data collection for the POI dataset and the corporate financial information. More detailed comments on the data are provided in the appendix.

---

<sup>13</sup>Originally I considered also the big box chains Walmart and Target but contrary to the neighborhood stores of Kroger and Safeway, I find that these chains are not operating in the same geographic markets or the markets would have to be defined extremely large such that assuming consumer to travel in such a large geographic area to purchase fresh grocery products becomes implausible in terms of irrational travel distances. The data indicate that only 67% of Walmart's stores are located in urban areas, while Target operates 81% of its stores in urban regions.

<sup>14</sup>POI stands for 'Point of Interest', an expression from the GPS technology where this datasets are used to provide GPS customers of any brand with an update of locations which might be of their interest when on the road. Some common examples of POIs other than supermarkets are hospitals, speed cameras or gas stations.

For the spatial analysis, I use the Geographic Information System ArcGIS. Based on free available polygon shape files for different spatial units in the U.S. with associated demographic characteristics, I define reasonable geographic markets and construct the following cross-section datasets:

$$Markets = \{Pop_m, HH_m, SQMI_m, N_m^K, N_m^S, Dist\_centroid_m^{DC,j}\}$$

$$Stores = \{X_s, Y_s, Diff_s^{comp}; Pop_{msj}, HH_{msj}, SQMI_{msj}, N_m^{own}, N_m^{comp}, Dist_{ms}^{DC,own}\}$$

The first dataset type consists of market level data. The observations are the markets, indexed by 'm', where at least one supermarket chain is active and associated variables like the number of stores per chain in each market ( $N_m^K$  for Kroger and  $N_m^S$  for Safeway respectively), the population and number of households per market ( $Pop_m, HH_m$ ), the geographic market size in square miles ( $SQMI_m$ ) and the distance from the market centroid to the closest regional distribution center of each firm ( $Dist\_centroid_m^{DC,j}$ ).

The second dataset consists of store level data and associated market data for a particular store  $s$ . The store data have been constructed using a vertical combination of the store dataset for each firm. The final dataset contains the projected store locations ( $X_s, Y_s$ ), the distance from the store to the closest regional DC of the respective chain affiliation ( $Dist_{ms}^{DC,own}$ ) and the distance to the closest competitor store within a market ( $Diff_s^{comp}$ ). The associated market features are as in the market level data.

### 3.2 Descriptive Proximity Analysis

In total Kroger counts 2.110 and Safeway 1.487 supermarket stores. Since I am interested in firms' location choice inside a market, I need to define reasonable shopping areas to identify where stores compete. In the literature based on Bresnahan and Reiss (1990) markets are usually defined as isolated cities. Recently Ellickson et al.(2011) proposed a variation where this assumption is relaxed allowing for market spillover effects for metropolitan and micropolitan areas but I find that this market definition is too large to be considered a shopping area for fresh grocery products. Instead I was looking for a market area definition such that consumers can be assumed to move in this area for grocery shopping given the data availability constraints of demographic and geographic market characteristics.<sup>15</sup> I propose 'Urban Areas' (UAs), densely settled census block groups that meet a minimum population density, as natural shopping areas for neighborhood supermarkets. To the best of my knowledge this definition has not been used so far in this context but the statistics show that this market definition captures almost all supermarkets in the data and yields reasonable travel dimensions for grocery products. I find that approximately 90 % of all neighborhood stores of the two considered chains are located in UAs which is taken as evidence for a natural shopping area for this type of stores. To illustrate where this markets are located, the appendix provides a map of the considered markets. Figure 6 provides a summary of the variables that will be used in the following analysis.

Considering the 'Contiguous U.S.', that is the United States excluding Alaska and Hawaii, Kroger, as the leading supermarket chain, operates in more markets than Safeway and

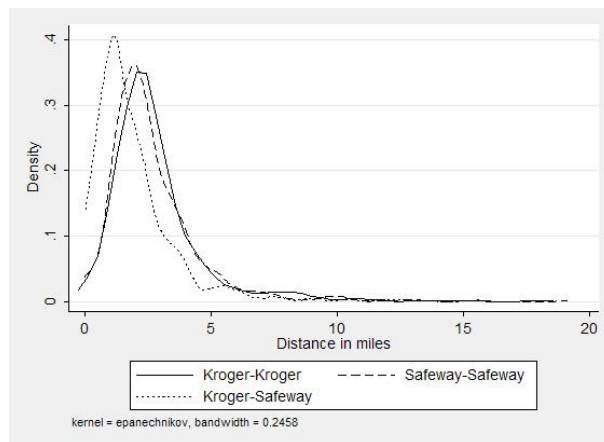
---

<sup>15</sup>This is in line with the geographic market definition by the European Commission which defines a retail market for daily consumer goods as "the boundaries of a territory where the outlets can be reached easily by consumers (radius of approximately 20 to 30 minutes driving time)" (COMP/M.5112 REWE/PLUS par.18, 2008).

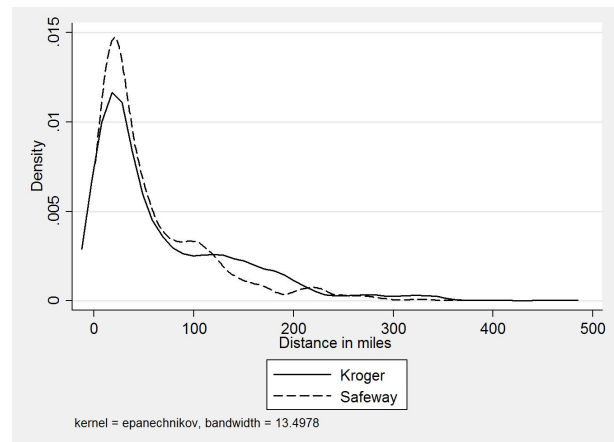
Figure 6: Summary Statistics

	Neighborhood supermarket chains	
	Kroger	Safeway
<b>Summary</b>		
▪ Total number of stores in UAs [%Total]	1.870 [89%]	1.373 [92%]
▪ Markets (UAs)	437	280
▪ Markets with at least 2 stores per chain	160	180
▪ Oligopolistic markets	67	67
▪ Total number of Regional DCs	25	12
<b>Proximity Measures</b>		
▪ E[Distance to closest own store]	2,82 (1,95)	2,68 (1,87)
▪ E[Distance to closest own store  Competition]	2,57 (2,03)	2,08 (1,97)
▪ E[Distance to closest competitor  Competition]	1,96 (1,89)	1,96 (1,89)
▪ E[Distance to closest RegDC]	68,59 (71,79)	57,20 (58,09)
▪ E[Distance to closest RegDC  Competition]	30,96 (42,33)	44,13 (51,30)
▪ E[number of stores in a market]	4,28 (12,70)	4,90 (13,67)
▪ E[number of stores in a market  Competition]	10,21 (24,53)	10,21 (20,97)
<b>Average Market Characteristics</b>		
▪ E[Population in a market]	183.742 (738.431)	337.681 (1.470.633)
▪ E[Number of Households in a market]	68.292 (256.462)	122.381 (525.735)
▪ E[Geographic market size (in sq. miles)]	72 (198)	94 (321)

Figure 7: Distance Distributions



(a) Differentiation



(b) Supply-Distance

counts more DCs, which is not surprising since being active in a larger geographical space, it's markets are organized in more distribution areas. However, the statistics show that the two main supermarket chains target similar markets. The data even suggest that their strategic entry decision is statistically equal in markets where they compete with each other (which I indicate in the statistics with the condition 'Competition').<sup>16</sup>

Section two of the summary statistics indicates proximity measures, that is the distances between stores of the same chain, to competitors and to the closest distribution center which supplies a certain chain. All the distances are measured in Euclidean distances in miles. Comparing the two main supermarket chains, an average distance of 2,82 or 2,68 miles respectively to own stores indicates similar pattern. Taking into account that Kroger indicates that it's supermarkets draw on average customers from 2,0-2,5 mile radius ('2011 Kroger Fact Book'), the average own-store-differentiation in the data suggests the existence of overlapping market areas. We will see if the supply consideration in the location choice partially explains this observations.<sup>17</sup> The average differentiation between competing chain stores in the same market is with 1,96 miles smaller than the differentiation to own stores. Figure 5 illustrates the distribution of the distance variables with more detail. We provide kernel density plots for the distance measures. Note that the own-store differentiation of both supermarket chains follow a very similar pattern. The same holds for the differentiation to the competitor in the markets where both chains are active but with a shift to the left which reflects the lower expected differentiation compared to own stores. Considering the distribution of the supply distance, both firms show a mode around 30 miles, from 80 miles until 100 miles and 120 miles respectively the distribution becomes almost flat and subsequently decreases. It reflects the colonization pattern of stores close to DCs while the flat region may suggest a kind of indifference belt followed by a possible distance threshold.

### 3.3 Empirical Analysis of Supply Distance Effects

In order to verify, whether the proposed mechanism suggested by the model yields a possible explanation for the observed pattern in the data, I run an empirical analysis using continuous distance measures. The variable of interest is the geographical differentiation between a store of chain  $i$  and a store of chain  $j$ , denoted as  $Diff^{comp}$ , which is estimated as a function of demand shifters ( $X$ ) and distribution aspects from the supply side ( $Y$ ).

$$E[Diff^{comp}|X, Y]$$

Recall that the constructed datasets of stores and markets, where at least one store per market is active, captures three possible market outcomes in terms of market presence of the supermarket chains. To verify the model, we are especially interested in firms' location choice under competition. To identify stores in competitive markets, I generate dummy variables depending on the number of stores that each firm operates in the market, in particular:

$$Competition = 1 \quad \text{if} \quad (N^K \geq 1, N^S \geq 1)$$

---

<sup>16</sup>I abstract from scale effects which might be larger for Kroger as the leading supermarket chain.

<sup>17</sup>Other reasons may be a pre-emptive behaviour of the firm, that is packing stores close together to foreclose the market, or agglomeration effects, which are not considered in this analysis.

Note that when analyzing the differentiation between stores of different chains, I implicitly select the competitive markets. However, since the selection rule  $s$  is a deterministic function of the market presence, which is captured in the matrix  $X$  as  $N_S$  and  $N_K$  respectively, we have  $E[u|X, S, s] = E[u|X, S]$  such that the selection issue can be ignored. So let's specify the model of geographic differentiation:

$$E[Diff^{comp}|X, Y] = \beta_0 + \beta_1 X + \beta_2 Y$$

$$\text{with } X = (Pop, SQMI, N^{own}, N^{comp})' \quad \text{and} \quad Y = (Dist^{DC, own}, (Dist^{DC, own})^2)'$$

The underlying intuition of this specification is based on the H&H model presented in the previous chapter. Note that if the DC costs were not considered in the firm's location choice, the differentiation should be independent of  $Y$ . I expect that if the distribution center is not too far away from the market, the stores consider the distribution distance in their location choice with respect to the competitor. However, when bringing the model to the data, I face two potential problems which are discussed in the following.

*Network problem.* The ideal experiment to analyze whether there is a distribution effect as specified in the model would be taking otherwise equal linear cities with  $N = 2$  stores each and random DC locations in space. However, contrary to the simplified theoretical model, in real world there are markets with more than two stores, that is a store network for which our linear model doesn't account for. In such markets, a supermarket has to consider the geographic differentiation to more than one competing store. It seems reasonable that the closest store, in terms of the Euclidean distance, matters most in the price competition but considering only the 'closest neighbor' ignores possible competition effects of other stores.<sup>18</sup>

*Simultaneity problem.* The aim is to explain the store-differentiation as a function of the location of the closest DC. If I use the distance to the closest DC as explanatory variable, I may introduce a simultaneity problem. If the distribution distance is endogeneously determined by the store's location choice, which is captured in the differentiation of the firms, the estimated coefficient  $\beta_2$  will be biased.

In a first step, to demonstrate the link between geographic differentiation and distribution distance, I run an ad-hoc analysis using the closest neighbor distances as endogenous variable and the store distance to the closest DC as explanatory variable.

In a second step, I address both of the mentioned problems at once using aggregated data. To address the network-problem, the easiest solution to implement is to redefine the dependent variable as the average differentiation within a market. This might be interpreted as a kind of representative differentiation within a market but comes at the cost of "loosing" observations when going from store-level data over to market-level data. For the analysis of the store differentiation to the closest competitor we are left with 67 observations (oligopolistic markets). For the purpose of this analysis I consider this small sample as still enough to get rid of the network problem at a low cost. A more sophisticated solution, using store-level data, would be to redefine nearness, taking the weighted average differentiation over an x-miles radius around each store or to set up a structural model.<sup>19</sup> For both solutions we need a detailed geography setup which goes beyond the

---

<sup>18</sup>A related problem arises for markets where A is the closest neighbor of B but for B the closest neighbor is C.

<sup>19</sup>For a discussion of "What is near?" see for example Miller(2004).



purpose of this paper. Hence, I implement the solution with aggregated data and leave the alternative option for further work. Moreover, the simultaneity problem is addressed with an IV approach. While the individual store distance to the closest DC may depend on the differentiation between stores, the distance of the exogenous DC location to the exogenous market centroid ( $Dist\_centroid^{DC}$ ) is supposed not to be correlated with the error term but is highly correlated with the store distance to the DC as well as the average supply distance within a market which yields a good instrument. Since I include level as well as squared distribution distances, I use both, the distance and the squared distance from the market centroid to the closest DC which are linearly independent instruments.

Figure 8: Regression results

	Store level – Coefficient estimates	
	(1) Kroger	(2) Safeway
Population (in 100T)	-0.02047***	-0.01015**
SQMI	0.00160***	0.00234***
$N_{Kroger} = \{1, 2, \dots, 10, 10+\}$	0.27359***	-0.56419***
$N_{Safeway} = \{1, 2, \dots, 10, 10+\}$	-0.36954***	0.62319***
$Dist^{DC}$ (in 100 miles)	-0.95976**	0.08477
$(Dist^{DC})^2$	0.34582**	0.00189
cons	2.42450***	1.332342***
observations	684	684
	Market level – Coefficient estimates	
	(3) Kroger	(4) Safeway
Population (in 100T)	0.01797	0.04977(*)
SQMI	0.00273**	0.00183(*)
$N_{Kroger} = \{1, 2, \dots, 10, 10+\}$	-0.01569	-0.07802***
$N_{Safeway} = \{1, 2, \dots, 10, 10+\}$	-0.04202**	.0217634
$E[Dist^{DC}]$ (in 100 miles)	-1.86810***	-1.87496***
$E[(Dist^{DC})^2]$	0.5346217***	0.67567***
cons	2.48834***	2.53548***
observations	67	67
***Significant at 1 percent level, **Significant at 5% level, *Significant at 10% level, (*) Significance at 12% by reason of small sample properties; Instrumented variables: $E[Dist^{DC}]$ , $E[(Dist^{DC})^2]$ , Instruments: $Dist\_centroid^{DC}$ , $(Dist\_centroid^{DC})^2$ and the respective explanatory variables.		

The regression table depicts the main results. Column (1)-(2) present the results of the ad hoc analysis based on store level data, with the geographic differentiation of Kroger (Safeway) from competing stores as dependent variable. It's worth mentioning that I include the number of stores, treated as exogenous in the location decision, since more stores in a small area necessarily cause a smaller store-differentiation. Not controlling for the number of stores would cause an omitted variable bias in the coefficient of the geographic area (square miles). Looking at the results, it attracts attention that for Kroger distribution costs are significant in the location choice while for Safeway they are not. Although, looking at the summary statistics, the firms seem to follow similar strategies,

the ad-hoc analysis suggests that only Kroger internalizes the cost effect. Column (1) suggest a stable U-shaped pattern of the distance to the DC. Together with the negative coefficient of the level distance and the positive intercept, the quadratic pattern implies that for 'small' supply distances the differentiation decreases in the supply distance while for large distances the differentiation increases. For Kroger, the minimum differentiation is reached at  $Dist^{DC,K} = 1.2$  (120 miles) which is in line with our conjecture about a kind of distance threshold when interpreting the distribution of the supply-distance in Figure 7. With respect to the model, the store-level results suggest that Kroger is playing a mixed-location strategy considering distribution costs in it's strategic positioning as outlined in the simple linear model with distribution costs. If the DC is more than 120 miles away, the firm differentiates more and more from it's competitor since distribution economies become less important which is in line with the extension of the model when  $z_2 > 0$ . Respective other demand shifter that effect the firm's location choice, the results are as expected. A larger market area provides more space for differentiation and also the number of stores for each chain restrictive for the location choice. Moreover an increase in the population decreases the differentiation between firms but we have to be careful with the interpretation since we don't control for the distribution of the consumers inside the market.

Considering the market-level regressions with instrumented distribution distances, column (3) and (4), the results confirm the U-shaped pattern for Kroger but also for Safeway the

being endowed ex-ante with a particular technology (incumbent product) and have to decide whether to develop it further in order to optimize their location in the product space.

Some particular examples may be found in the software or automobile industry, both industries with labour-intensive and complex development processes that require specialized skills. For instance, for software vendors, it may be more efficient to sell relative generic software packages at competitive prices rather than more specialized software solutions that imply high development costs. In the automobile industry, we may think of a particular car model of each manufacturer and their decision about the new generation of the cars, that is how far away the engineers move from the characteristics of the original model. In other words, if the fixed R&D costs are internalized in the product design decision, the cost consideration can change the optimal location in the product space relative to a pure demand-based decision.

## 5 Conclusion

I have provided a theoretical model and empirical evidence how the consideration of operational efficiency, in terms of supply costs, in the firms' optimal location choice affects the degree of product differentiation among firms. The proposed model has shown that accounting for firms' distribution costs in an otherwise standard Hotelling framework, maximal horizontal differentiation of competing stores might no longer be optimal. We have provided conditions under which the trade-off between demand and cost considerations in the firms' strategic location choice induces an optimal location in the interior of the market. Although firms earn less marginal revenues due to the increased price competition, in terms of net profits they are better off than ignoring distribution economies and treating supply costs as exogenous once they are established. But also consumers benefit from the hybrid location strategy of the firms since they face lower prices and spend an aggregate amount of transportation costs which is less (or equal) than in the standard model. The empirical verification of the model for optimal supermarket locations suggests that supermarket chains consider distribution distances in their location choice. The optimal degree of geographic differentiation to the competitor depending on the distance to the closest distribution center is U-shaped, declining for small or moderate distribution distances and increasing for long distances. The result is in line with the theoretical model suggesting that a mixed-location strategy is profit-maximizing.

Theory and empirics suggest that the trade-off between competition effect and distribution economies is the strongest when the distribution facility is relatively close to the market where the stores are operating. If the DC is too far away the distribution economies decrease and the competition effect dominates the degree of product differentiation.

This paper is a first step for a better understanding of firms' optimal location choice in distribution intensive industries and provides likewise incentives for further research on alternative interpretations of the extended Hotelling model.

## References

- [1] Andersen, Michael M. and Poulsen, F. (2006). Discount Business Strategy: How the new market leaders are redefining business strategy. *Wiley*.
- [2] Anderson, A., de Palma, Thisse, J.F. (1992). Discrete Choice Theory of Product Differentiation. *MIT Press, Cambridge*.
- [3] Anderson, S. (1988). Equilibrium existence in the linear model of spatial competition. *Economica*, Vol. 55, pp.479-491.
- [4] Bresnahan, Timothy F. and Reiss, Peter C. (1990) Entry in Monopoly Markets, *The Review of Economic Studies Ltd.*, Vol.57, No.4, pp. 531-553.
- [5] D'Aspremont, C., J. Jaskold Gabszewicz, J.F. Thisse (1979). On Hotelling's 'Stability in competition'. *Econometrica*, Vol. 47, No. 5, pp.1145-1150.
- [6] Datta, Sumon and Sudhir, K. (2011). The Agglomeration-Differentiation Tradeoff in Spatial Location Choice. *Working Paper*.
- [7] Ellickson, P.B., Houghton, S. and Timmins, C. (2010). Estimating Network Economies in Retail Chains: A Revealed Preference Approach. *Working Paper*.
- [8] Hamoudi, Hamid and Moral, Mara J. (2005). Equilibrium Existence in the Linear Model: Concave versus Convex transportation costs. *Papers and Regional Science*, 82, pp.201-219.
- [9] Holmes, Thomas J. (2011). The Diffusion of Wal-Mart and Economies of Density. *Econometrica*, Vol. 79, No. 1, pp. 253-302.
- [10] Hotelling, H. (1929). Stability in Competition. *The Economic Journal*, 39, pp. 41-57.
- [11] Matsa, David A. (2011). Competition and Product Quality in the Supermarket Industry. *Quarterly Journal of Economics*, Vol. 126, No. 3, pp.1539-1591.
- [12] Mazzeo, Michael J. (2002). Product choice and oligopoly market structure, *RAND Journal of Economics*, Vol.33,No.2, pp. 221-242.
- [13] Meagher, Kieron J., Teo, Ernie G.S. and Wang, Wen (2008). A duopoly location toolkit: Consumer densities which yield unique spatial duopoly equilibria. *The B.E. Journal of Theoretical Economics*, Vol. 8, No. 1, Article 14.
- [14] Miller, Harvey J. (2004). Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers*, Vol.94, No. 2, pp.284-289-
- [15] Seim, Katja (2006). An empirical model of firm entry with endogenous product-type choices. *RAND Journal of Economics*, Vol. 37, No. 3, pp.619-640.
- [16] Tirole, J. (1998). The theory of Industrial Organization. *MIT Press, Cambridge*.
- [17] Thomadsen, Raphael (2007). Product Positioning and Competition: The Role of Location in the Fast Food Industry. *Marketing Science*. Vol.26, No. 6, pp.792-804.

## A Proofs and Algebraic Details

**Symmetric solution of location choice.** We can either see it directly from the best response function or solve for it analytically if we subtract  $BR_a^1 - BR_b^1$  and solve the quadratic equation under the feasibility constraint:

$$(a^2 - b^2) * \left(-\frac{2}{9}t\right) + (a - b) \left(-\frac{1}{3}t - \frac{1}{9}tr - 2d_1\right) \leq 2d_1 (r - z_1^b - z_1^a)$$

Define  $\gamma \equiv \frac{t}{d}$  and  $\bar{z} = \frac{z_1^a + z_1^b}{2}$ , then

$$a(b) = \begin{cases} b & \text{if } z_1^a = r - z_1^b \ (\Leftrightarrow \bar{z}_1 = \frac{1}{2}), \\ -1 - 9/(2\gamma) + \sqrt{[b + 1 + 9/(2\gamma)]^2 + (9/\gamma)(2\bar{z}_1 - r)} & \text{otherwise} \end{cases}$$

In this paper I focus on the symmetric location equilibrium but I may conjecture that whenever the symmetry condition  $z^a = r - z^b$  doesn't hold, there exists an asymmetric location equilibrium iff the cost advantage of the market leader is not too strong.

**General DC location.** Under symmetry, the optimal location is implicitly given by  $F = a \left(-\frac{2}{3}t - 2d_1\right) + d_2 \cdot \frac{1}{\sqrt{1 + \left(\frac{z_2}{z_1 - a}\right)^2}} - \frac{1}{6}tr + 2d_1z_1 \leq 0$ . Since  $F(a = -\infty) = -\infty$  and

$F(a = +\infty) = +\infty$  and  $F$  is continuous in  $a$ , there is at least one root. And since all the summands of  $F$  are monotonic on the interval  $[0, r]$  or a constant, the root is also unique. To check whether there is an interior solution on the Hotelling Line it is enough to check for a positive root which is the case whenever  $F(0) > 0$ . Evaluating  $F$  at zero yields the following condition for an interior solution:  $\frac{d_2z_2}{\sqrt{z_1 + z_2}} + 2d_1z_1 \leq \frac{1}{6}tr$ .

**Proof Proposition 1.** The firm's best location is an interior solution on the Hotelling Line if the consumers' travel cost parameter is small enough (relative to the distribution cost parameter) such that  $t < t_{crit}(b) = \left[2d_1z_1^a + \frac{d_2z_1^a}{\sqrt{(z_1^a)^2 + (z_2^a)^2}}\right] \left[-\frac{1}{18}b^2 + \left(\frac{1}{6} - \frac{1}{18}r\right)b + \frac{1}{6}r\right]^{-1}$ .

Under symmetry, the condition collapses to  $t < t_{crit} = \frac{1}{r} \left(12d_1z_1^a + \frac{6d_2z_1^a}{\sqrt{(z_1^a)^2 + (z_2^a)^2}}\right)$ .

Since for  $z_1 > a$  the MRDE(a) are strictly increasing in  $a$  and MRPD(a) are strictly decreasing in  $a$ , if  $MRDE(a = 0) > MRPD(a = 0)$  the firm chooses a corner solution,  $a = 0$ . From the equilibrium condition (1) it can be seen that this is the case whenever  $\left(\frac{1}{18}t\right)b^2 + \left(\frac{1}{18}tr - \frac{1}{6}t\right)b - \frac{1}{6}tr < -2d_1z_1 - \frac{d_2z_1}{\sqrt{z_1^2 + z_2^2}}$  and under rearrangement we can establish a critical value  $t_{crit}(b; z, t, d)$ .

If  $t > t_{crit}$ , the demand effect dominates the supply effect and the firm finds it optimal to choose maximal differentiation. Note that if  $z_2 = 0$ ,  $d_2 = 0$  or  $d_1 = d_2$ , I could define a relative threshold  $(t/d_1)_{crit}$ . Under symmetry  $a = b$ , the threshold reduces to  $t_{crit} = \frac{1}{r} \left(12d_1z_1 + \frac{6d_2z_1}{\sqrt{z_1^2 + z_2^2}}\right)$ .

**Proof of Proposition 2.** First, recall that the effect of the market distance from the exogenous DC location is linear separable from the market events, that is the hypothetical case that agents as well as DC(s) are located inside the market. Hence, I take the derivative of the *scale-effect* with respect to the distance between market and DC location ( $z_2$ ) :

$$\frac{\partial^2 DC_a}{\partial a \partial z_2^a} = \frac{d_2(z_1^a - a)z_2^a}{[(z_1^a - a)^2 + (z_2^a)^2]^{\frac{2}{3}}} \geq 0 \quad \text{for } z_1^a \geq a$$

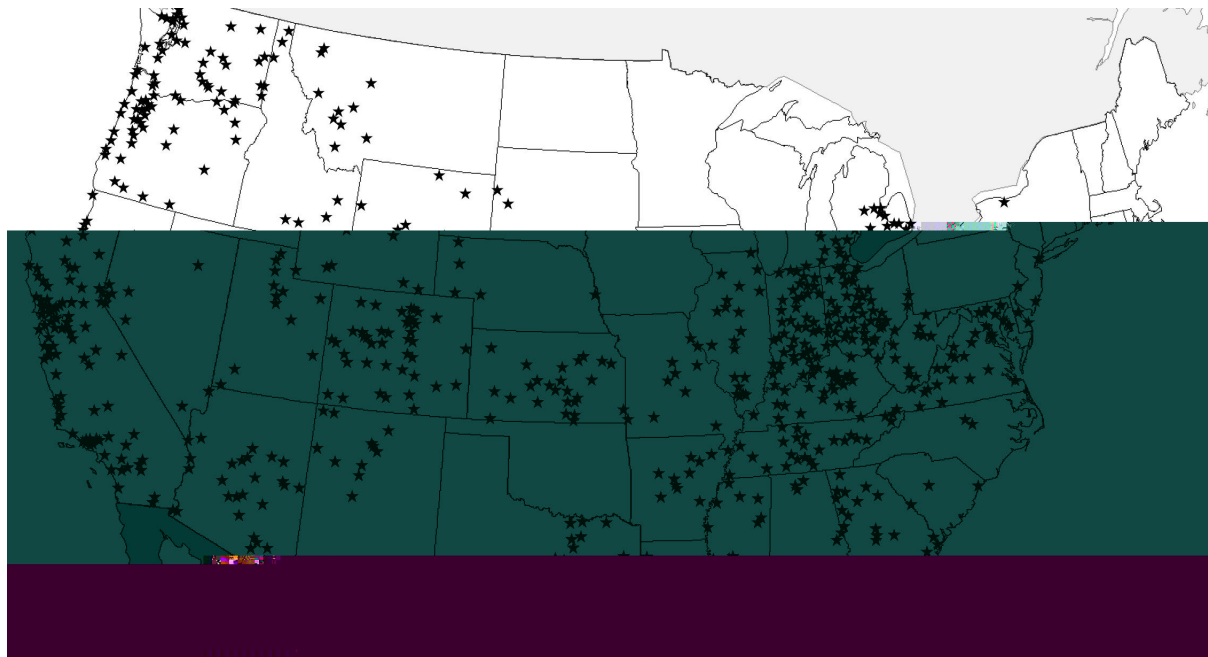
Since the MRDE are negative for any  $z_1^a > a$  (indicating marginal cost savings), the positive sign of the second derivative implies diminishing distribution economies as the distance of the DC to the market becomes large.

## B Detailed explanation of the data

**Kroger.** To identify supermarkets which are operated by The Kroger Company, I use a free POI file from July 2012 identifying the geographic coordinates and banners for all grocery stores which are under the firm's ownership ([www.poi-factory.com](http://www.poi-factory.com)). Additional information from the firm's web site allows to identify the store format which is operated under each banner. ([www.thekrogerco.com](http://www.thekrogerco.com)). The GPS data provide a total number of 2,428 grocery retail stores in the U.S. of which 2,110 are supermarkets, 146 are warehouse stores and 172 are multi-department stores (similar to supercenters). The data are consistent with the firm's public information indicating in may 2012 a total number of 2,425 grocery retail stores, that is 3 stores less than the data which I assign to the two month difference between this data sources. (The data consistency holds also for the firm's convenience stores, which differ in only 3 stores with 786 stores registered in the POI dataset and 789 stores indicated by the firm in may 2012.) The locations of the distribution facilities are collected from the firm's 'Ship-to Warehouse Location List' for vendors who are required to use EDI (Electronic Data Interchange). In 2012 the warehouse location list indicates 34 distribution divisions of which 27 are local distribution divisions and 7 are supraregional consolidation warehouses, denominated 'Peyton's DC' and 'Goddard Western DC'. While some divisions have only one big local distribution center others have several specialized warehouses located next to each other, in the latter case I took the street address of the most general one for the geocodification. The information is consistent with other firm sources like the '2011 Fact Book', which indicates 34 distribution centers. It is worth mentioning that some DCs are operated by the firm itself while others are operated by third party service providers, which is as a result of Kroger's outsourcing and remodeling of its distribution network during the last years. When analyzing the subsample of the supermarket format only, I exclude the FredMeyer Regional DC (division 22) which supplies the multi-department stores that are operated under this banner.

**Safeway.** To identify stores and DCs of Safeway I use two types of sources. First, I use a POI dataset which identifies all Safeway facilities in the U.S. and Canada based on firm information. The data set provides locations for all stores of any brand as well as associated distribution centers which were operated in March 2008. After sorting out the number of retail stores in the US we are left with 1,545 store locations in the U.S., of which 973 are operated under the Safeway banner, 300 Vons, 116 Randalls, 80 Dominick's, 37 Genuardi's and 39 Carrs (I eliminate one observation 'Citrine Bistro'). A comparison with data from the 'SW Fact Book 2008' and the '2007 Annual Report', show an acceptable difference of 24 stores. The US stores are assigned to 9 Operation Areas (Divisions) which are supplied by 13 main distribution centers. In general each division has one regional DC, exceptions are South California (Vons) and Texas (Randalls) which have two DCs each and Seattle which is supplied by even three different DCs. Complementary information from the firm's web site allows to match each store with the corresponding distribution center by the division.

**Market definition.** Markets for 'neighborhood grocery stores' are defined as urbanized areas (UAs). We have shown that this particular definition is convenient for the Kroger-Safeway-data. To illustrate where this markets are located, the map below indicates all the UAs where at least one of the firms is present.



Considered markets (UAs) in the US with active Kroger-stores and/or Safeway-stores