| | |
|---|---|
| File ID | 104238 |
| Filename | 5 Competition and entry in a spatial model with applications to banking |

SOURCE (OR PART OF THE FOLLOWING SOURCE):

| | |
|---|---|
| Type | Dissertation |
| Title | Essays on bank monitoring, regulation and competition |
| Author | M. Marinc |
| Faculty | Faculty of Economics and Business |
| Year | 2008 |
| Pages | 226 |
| ISBN | 9789051708851 |

FULL BIBLIOGRAPHIC DETAILS:
http://dare.uva.nl/record/270596

# Chapter 5

# Competition and Entry in a Spatial Model with Applications to Banking


**Abstract**

This chapter recasts the analysis of Chapter 3 in a Hotelling framework. It analyzes entry in a spatial model in which firms are heterogeneous in their production costs. The key result is that intensifying competition by lowering transportation costs can augment expected profits and entry, essentially by shifting the market share from high-cost to low-cost firms. It is also shown that increasing the differences in firm costs augments firm profits, entry, and social welfare. Lowering production cost of high-cost firms augments entry if competition is low (i.e., for high transportation costs) but hampers entry if competition is high (i.e., for low transportation costs). When specifically applied to banking, the results of the analysis show that *i*) deposit insurance may lead to less entry, *ii*) capital regulation may lead to more entry in banking, and *iii*) bailout policies may reduce bank profits and entry, even if they come at no additional cost for banks.



Keywords: Entry, Spatial Competition, Transportation Costs, Banking
JEL CLASSIFICATION: L11, G21

## 5.1 Introduction

The perceived distances between firms and consumers has decreased substantially due to developments in IT and the lifting of regulatory barriers. Firms operate in increasingly competitive and dynamic environments in which impediments to cross-border entry are low. In this chapter I pose two questions: *i)* how does the level of interfirm competition in the market (measured by the level of transportation costs) affect a firm's entry decision in such a market; and *ii)* how do the changes in firm heterogeneity affect firm profits, entry, and social welfare.

I analyze these issues in a spatial monopolistic competition setting which distinguishes two types of firms, that differ in their cost of production. In particular, high-cost firms incur higher per unit production costs than low-cost firms. Firms are allowed to compete for consumers (located uniformly on a unit circle). The level of competition is determined by the size of transportation costs that consumers incur when traveling to the firms. This chapter recasts the analysis in Chapter 3 in a Hotelling framework. While the analysis has generic interest in most of the application, I will look at the banking industry; see Degryse and Ongena (2005) for the relevance of "distance" to bank competition.

The key result is that increasing competition in the market through lowering transportation costs might be beneficial for entering firms and could bring more entry essentially by redistributing a market share from high-cost to low-cost firms. The shift in the market share could augment the expected profits of entering firms. Although the redistribution of the market share that lowering transaction costs entails is beneficial for entering firms, the net effect on entry could go either way because competition also transfers rents from firms to consumers. I establish conditions for which competition encourages entry. Roughly speaking, higher competition in terms of lower transportation costs has a positive effect on entry provided there are sufficiently many (but not too many) low-cost firms in the economy *and* competition is already high (i.e., entry costs should be sufficiently low). I show that for such intermediate quality industries increasing competition reduces the competitive strength of high-cost firms and also encourages entry.

Salop (1979) argues that lowering transportation costs increases competition and hampers firm profits. This then results in a lower number of entering firms. However, Salop shows that this intuitive result reverses for high transportation costs (i.e., at lower levels of competition). In particular, a "perverse" equilibrium may exist, in which lowering transportation costs leads to higher profits and more entry. This is because (almost) monopolistic firms spend less to compensate consumers on transportation costs. In particular, high transportation costs prevent most of consumers to travel to a competing firm. The firm may act (almost) as a monopolist and may charge to consumers (most of) what they are willing to pay for the products. If transportation costs decrease, consumers are willing to pay more for the same product knowing that they incur lower traveling costs when buying it. This underscores that entry is a non-monotonic function of transportation costs in homogeneous industries (i.e., industries in which firms have equal production costs). More specifically, entry increases in transportation costs for low transportation costs but decreases in transportation costs for high transportation costs.

This approach deviates from the extant spatial competition literature in that it recognizes differences in cost efficiency among firms.[1] Allowing for heterogeneity (i.e., different production costs) among firms may again reverse the standard result in Salop (1979). Another "perverse" equilibrium may appear, this time for low transportation costs (for high levels of competition). More specifically, if competition is fierce, a decrease in transportation costs may augment entry. The rationale for this is the following. Higher competition improves the total efficiency of the industry through the shift of market share from high-cost to low-cost firms. This then results in higher firm profits and higher entry.[2] If competition is low, however, efficiency increases only slightly and the standard Salop effect is at work; namely, profits go down due to the negative effect of competition on firm rents. In sum, entry is a U-shaped function of the level of transportation costs; that is, entry decreases in transportation costs for low transportation costs and increases in transportation costs for high transportation costs.

Melitz (2003) analyzes the intra-industry effects of international trade using a Dixit and Stiglitz (1977) model of competition extended for firm heterogeneity. He shows that opening up borders for trade increases the total productivity of firms, essentially by a reallocation of resources from low-quality to high-quality firms. In his model, firms first enter the industry and later enter export markets. He shows that a decrease in trade costs generates entry of new firms into the export markets. At the same time, firms exit the industry. He does not offer an answer for how trade costs influence the total number of firms in the industry. The main analysis here complements Melitz's research by showing that *initial entry* into the industry may increase and consequently the total number of firms may go up with lower costs of trade.

I also analyze the effect of firm heterogeneity on firm profits, entry, and social welfare. I show that increasing the differences in costs between firms may be valuable for firms even though it lowers their cost efficiency. The reason for this is that higher differences between firms lower the level of competition. A firm's profits may depend only on its relative advantage to its competitor. More specifically, the bigger the differences in firm costs, the higher rents firms can extract from their consumers. Whereas lower cost efficiency may be passed on to consumers, increased differences between firms augment firm profits and entry. Even more, increasing the cost differences between firms may also improve social welfare. That is, competition increases the efficiency of the industry through the shift of the market share from high-cost to low-cost firms. For more heterogeneous firms, a redistribution of market share is more pronounced and, consequently, social welfare increases.

I further analyze the role of cost differences to show that increasing the competitive

[1] Empirical work shows that the magnitude of variation in cost efficiency at the plant-level is enormous and persistent even within narrowly defined industries (see Bartelsman and Doms (2000)).

[2] Recent literature that discusses the positive relation between competition and productivity includes Aghion and Howitt (1992), Nickell (1996), and Raith (2003).

ability of low-cost firms always augments entry, whereas enhancing the competitive ability of high-cost firms might hamper entry. That is, improving the cost efficiency of a low-cost firm augments the average firm rents and also increases the differences between firms. Both effects augment profits and entry. However, enhancing the cost efficiency of a high-cost firm lowers the differences in costs between firms and hampers the shift of market share from high-cost to low-cost firms. If competition is high, and the industry is of high quality (i.e., there are many low-cost firms), this results in decreased profits and entry.

I apply the results to banking. More specifically, I provide a spatial setting for the analysis of Chapter 3, which was derived in the context of capital regulation in banking. Chapter 3 models bank competition in a simplified search model. This chapter shows that similar results can also be derived in a spatial context.[3] Whereas Chapter 3 exclusively focuses on the competition effects of capital regulation, I focus here on a broader set of regulatory practices in banking. I first explore the impact of branching deregulation on entry in banking. Interestingly, banks on average may have incentives to push for branching deregulation. Branching deregulation brings beneficial redistribution of market share to better banks. In spite of higher competitive pressure, this could improve profitability of those banks.

I also show that policies enhancing market disclosure and market pressure should be the most valuable in highly competitive banking industries. That is, market disclosure and market pressure enhance the differences between banks. This is especially valuable if competition between banks is high. That is, competition shifts the market share to the most efficient banks. This leads to higher profits and augments social welfare.

Deposit insurance decreases the differences between banks as it mostly helps weak banks. I show that in highly competitive, high-quality banking systems banks on average may lobby for limits on deposit insurance. To the contrary, in less competitive banking markets banks may prefer to have more generous deposit insurance.

I also argue that extensive bailout policies augment entry in banking if competition is low but hamper entry if competition is high. Bailout policies essentially lower the costs of financing for banks. Especially, high-cost (riskier) banks obtain funds at better terms than without bailout policies. If competition is low, this increases bank profits and augments entry. However, in the highly competitive banking industry increasing the competitive strength of riskier banks has a negative effect on entry. More specifically, bailout policies unfairly help high-cost (risky) banks and this is detrimental for low-cost (safe) banks. Low-cost banks cannot compete on a fair ground with high-cost banks, which lowers expected profits and hampers entry. Consequently, the banking industry as a whole may push for more stringent bailout policies.

The simplicity of this model does not come without some concessions. First, in this model

[3]A spatial model may also literally apply to banking, where geographical distances between banks and borrowers could play a role. For example, Degryse and Ongena (2005) show that distances between borrowers and their banks, (as well as distances between borrowers and competing banks) play an important role in the Belgian banking system.

entering firms are not aware of their types. In reality one would expect entering firms to have at least some prior knowledge about their costs. Second, firms in this model incur an entry cost only when entering. Later, this cost is sunk and high-cost firms do not exit the market. However, this aligns with empirical research that indicates that newly entering firms have higher failure rates than incumbents (see Dunne, Roberts, and Samuelson (1989)). This suggests that entering firms first make irreversible investments in production and only later discover their productivity (their type).

The chapter is organized as follows. In Section 5.2 I develop the model. Section 5.3 contains some basic analyses. In Section 5.4 I present the main findings. I first discuss the model with homogeneous firms. Later I extend it to heterogeneous firms. Section 5.5 contains the application to banking. Section 5.6 concludes this chapter.

## 5.2 The Model

### 5.2.1 Preliminaries

There are two players in the model: firms and consumers. Firms produce products that consumers buy. Firms differ in the variable cost of production. In particular, low-cost firms have low per-unit costs of production whereas high-cost firms incur high production costs.

I describe competition with a spatial model of price discrimination. Consumers are located on a unit circle. Firms compete for each consumer by setting a price that depends on the consumer's location. That is, firms can perfectly price discriminate between differently located consumers which is a typical situation in banking (see Degryse and Ongena (2004) (2006), Petersen and Rajan (2002) and Agarwal and Hauswald (2007)). The results could also be replicated in the uniform pricing model (as used by Salop (1979)) in which firms would set one price for all consumers regardless of their location.

The consumer decides on a firm by comparing price offers and firm locations (and corresponding transportation costs). In particular, a consumer located further away from the firm is prepared to pay less for its product because of two reasons. First, he has to overcome higher transportation costs to reach the firm. Second, he is located closer to a competing firm, which makes an offer of a competing firm more attractive.

The level of competition is determined by the level of transportation costs. If transportation costs are high, traveling to the distant firm might be prohibitively expensive and consumers might be locked in to the closest firm. For lower transportation costs, consumers are more willing to consider an offer of a more distant firm. Consequently, firms extract lower rents from consumers. This horizontal differentiation of firms due to transportation costs is combined with vertical differentiation determined by differences in firms' production costs.

$t = 0$:

♠ Firms enter the industry at a cost $F$.

♠ Firms distribute themselves uniformly on the unit circle.

$t = 1$:

♠ Firms discover their types.

♠ Each firm offers a distinct price to each consumer.

♠ Each consumer chooses a firm to purchase a product from.

♠ Firms produce and consumers make their purchases.

Figure 5.1: Timeline

### 5.2.2 Model details

*Timing and information structure:* There is universal risk neutrality, with a discounting factor normalized to 1. There are two dates, $t = 0$ and $t = 1$. At $t = 0$, firms enter the industry at a cost $F$. Upon entering, a firm does not know its type. Firms locate symmetrically on the unit circle. At $t = 1$, firms and consumers discover other firms' types. Consumers then decide from which firm they will purchase a product. The chosen firm commences with production. Finally, each consumer travels to the chosen firm to collect the product. Figure 5.1 summarizes the sequence of events.

*Firms:* Firms choose to enter the industry at $t = 0$. All firms are initially (perceived as) identical. Entering $N$ firms distribute symmetrically on the unit circle.[4] At $t = 1$, firms learn whether they are high-cost ($B$) or low-cost ($G$) firms. Low-cost firms have a lower variable cost of production $c_G$ than high-cost firms $c_B$; that is, $\Delta c \equiv c_B - c_G > 0$. The cross-sectional probabilities of having high costs ($\gamma$) or low costs ($1 - \gamma$) are known to all.

*Consumers:* Consumers are uniformly distributed on the unit circle. Their total mass is normalized to one unit. A consumer has inelastic demand and purchases 1 unit of a product if its price is lower than the reservation price $Y$. The effective price for a product equals the factory door price set by the producer plus a transportation cost of $\tau$ per unit of distance that the consumer incurs when purchasing the product.

*Competitive Environment:* Competition between firms occurs at $t = 1$ when the types of the firms are already public knowledge. Each firm makes a (potentially different) price offer to each consumer. That is, firms can price discriminate between consumers on the basis of their locations. Each consumer compares price offers of firms with their distances and accepts the best offer. The chosen firm then starts producing. Consumers travel to the chosen firm, incurring transportation costs, and buy the product (see also Figure 5.1).

[4]Because all firms are perceived as identical at that moment, this even distribution of consumers over all firms is quite natural. Bhaskar and To (2004) show that firms locate symmetrically in the circular model with perfect price discrimination. In the uniform pricing model, Lederer and Hurter (1986) show that, although firms never locate coincidentally in the Hotelling model, they locate symmetrically on the circular model.

## 5.3 Some Preliminaries

I solve the model through backward induction. First, I compute a firm's pricing decision at $t = 1$, holding the number of competing firms fixed at $N$. Subsequently, I allow for entry. Last, I compute social welfare.

### 5.3.1 Price offers and profits

Firms are located symmetrically with a distance $\frac{1}{N}$ between adjacent firms. Competition between two adjacent firms of type $i$ and type $j$ located at 0 and $\frac{1}{N}$ that set prices $p_i$ and $p_j$ evolves as follows. A consumer that is separated at distance $x$ from a firm of type $i$ and $\frac{1}{N} - x$ from an adjacent firm of type $j$ is indifferent between the offers of both firms if

$$Y - p_i - \tau x = Y - p_j - \tau[\frac{1}{N} - x]. \tag{5.1}$$

The lowest price that firms are prepared to offer equals their variable costs. The most distant consumer that a firm of type $i$ can obtain when bidding against a firm of type $j$ is at a distance $\bar{x}$, where $\bar{x}$ follows from (5.1); that is, substitute $p_i = c_i$ and $p_j = c_j$ in (5.1) to obtain

$$\bar{x} = \frac{c_j - c_i}{2\tau} + \frac{1}{2N}. \tag{5.2}$$

I now make the first assumption (recall that $\Delta c = c_B - c_G$).

**Assumption 5.1.**: $\Delta c < \frac{\tau}{N}$.

This assumption guarantees that the difference in cost efficiency between firms is small enough such that even high-cost firms obtain some market share. More specifically, a high-cost firm offers at most $Y - c_B$, whereas an adjacent low-cost firm could offer more, that is, $Y - c_G$. However, Assumption 5.1 guarantees that a high-cost firm can attract at least the closest consumer because of its proximity advantage; that is, $Y - c_B > Y - c_G - \frac{\tau}{N}$. Hence, the consumer in the location of a high-cost firm does not switch to an adjacent low-cost firm. This also implies that each firm competes only for the consumers located between itself and an adjacent firm, but not for consumers on the other side of an adjacent firm.

Now I compute optimal price schedules. I separate two cases. First, if competition is high (i.e., $\frac{\tau}{N} \leq Y - c_j$), a firm of type $j$ can reach all consumers between itself and an adjacent firm. That is, a firm of type $i$ can make an acceptable offer conditional on the adjacent firm not bidding. In particular, the right side of (5.1) is positive for $p_j = c_j$ and for $x \in (0, \frac{1}{N})$. A firm of type $i$ uses its advantage in transportation costs such that it charges price $p_i(x)$ for consumers located at the distance $x \in (0, \bar{x})$. I then insert $p_j = c_j$ into (5.1) and compute for $p_i$ to obtain

$$p_i(x) \quad = \quad c_j + \frac{\tau}{N} - 2\tau x. \tag{5.3}$$

A consumer at $x$ is indifferent between a firm of type $i$ offering $p(x)$ and the best offer that an adjacent firm of type $j$ can make, (in such a case, the consumer decides on a firm of type $i$).

Second, if competition is low (i.e., $\frac{\tau}{N} > Y - c_j$), an adjacent firm of type $j$ cannot reach the most distant consumers located closest to a firm of type $i$. We define

$$\underline{x} = \frac{1}{N} - \frac{Y - c_j}{\tau}. \tag{5.4}$$

Consumers located very close to a firm of type $i$ (i.e., $x < \underline{x}$) are located too far from an adjacent firm of type $j$ to buy its product regardless of its price. That is, the transportation cost $[\frac{1}{N} - x]\tau$ to reach a firm of type $j$ is greater than the consumer's rent $Y - p_j$ even if it charges the lowest price $p_j = c_j$. Those consumers cannot switch firms; hence they accept the monopolistic price of the firm of type $i$. In particular, the firm charges the reservation price of the consumer $Y$ lowered by $\tau x$ to compensate consumers for their transportation costs. The price schedule is

$$\hat{p}_i(x) = \begin{cases} c_j + \frac{\tau}{N} - 2\tau x & \text{if } x > \underline{x}, \\ Y - \tau x & \text{if } x \leq \underline{x}. \end{cases} \tag{5.5}$$

I now make the second assumption.

**Assumption 5.2.**: $Y - c_B - \frac{\tau}{2N} > 0$.

Assumption 5.2 guarantees that the reservation value of the product is high enough such that the consumer in the middle between two adjacent high-cost firms is prepared to purchase a product. This guarantees that the entire market is covered and that two adjacent firms compete for at least one consumer in the middle between them.

Now I compute firm profits. I again consider two cases. For $\frac{\tau}{N} \leq Y - c_j$, the price schedule in (5.3) applies and the expected profit of a firm of type $i$, competing with two adjacent firms of types $j$, is

$$\Pi_{i,j}(N) = \int_{-\bar{x}}^{\bar{x}} [p_i(x) - c_i] \mathrm{d}x, \tag{5.6}$$

where $\bar{x}$ is as given in (5.2). For $\frac{\tau}{N} > Y - c_j$, the price schedule in (5.5) applies and the expected profit is

$$\Pi_{i,j}(N) = \int_{-\bar{x}}^{\bar{x}} [\hat{p}_i(x) - c_i] \mathrm{d}x. \tag{5.7}$$

Inserting (5.3) into (5.6) and (5.5) into (5.7) yields[5]

$$\Pi_{i,j}(N) = \begin{cases} \frac{\{\tau+N[c_j-c_i]\}^2}{2\tau N^2} & \text{if } \frac{\tau}{N} \leq Y - c_j, \\ \frac{2Y-c_j-c_i}{N} - \frac{\tau}{2N^2} + \frac{[c_j-c_i]^2-2[Y-c_j]^2}{2\tau} & \text{if } \frac{\tau}{N} > Y - c_j. \end{cases} \tag{5.8}$$

## 5.3.2 Endogeneous entry

We now endogenize $N$, and hence allow for entry. The entry decision is made at $t = 0$. At that moment, each prospective firm does not yet know its own (future) type, but assesses its expected production costs based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$ of being a low-cost or high-cost firm. Each firm computes whether its expected profits from entering exceed the cost of entry F, anticipating the competitive environment (including the number of firms already present).[6]

The expected firm's profit after entering (i.e., after the entry cost $F$ has already been incurred) is

$$E(\Pi(N)) = \gamma^2\Pi_{G,G} + \gamma[1-\gamma][\Pi_{G,B} + \Pi_{B,G}] + [1-\gamma]^2\Pi_{B,B}. \tag{5.9}$$

I can now prove the following lemma (insert (5.8) into (5.9)).

**Lemma 5.1.** *The expected profit of an entering firm is*

$$E(\Pi) = \begin{cases} \frac{\tau}{2N^2} + \frac{\gamma[1-\gamma]\Delta c^2}{\tau} & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ \frac{2[Y-E(c)]}{N} - \frac{\tau}{2N^2} - \frac{[Y-E(c)]^2}{\tau} & \text{if } \frac{\tau}{N} > Y - c_G, \\ \gamma\frac{\tau}{2N^2} + [1-\gamma]\{\frac{2[Y-c_B]}{N} - \frac{\tau}{2N^2} + \frac{\gamma\Delta c^2-[Y-c_B]^2}{\tau}\} & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G, \end{cases} \tag{5.10}$$

*where $E(c)$ is defined as the expected production cost, that is $E(c) = \gamma c_G + [1-\gamma]c_B$.*

Lemma 5.1 shows that three regions of competition are distinguished. In the most competitive region, ($\tau \leq [Y - c_B]N$), adjacent firms can attract all consumers between them, even if firms are of high-cost type. In the least competitive region ($\tau > [Y - c_G]N$), adjacent firms act as monopolists for at least some of the closest consumers even if they are of low-cost type. The region with intermediate competition ($Y - c_B < \frac{\tau}{N} \leq Y - c_G$) provides the most fruitful results. Here low-cost firms can reach all consumers, but high-cost firms cannot reach the most distant ones. That is, if a firm competes with a low-cost firm, it must compete for

[5]Thisse and Vives (1988) similarly observe two different price schedules. In my model, the demand for each consumer is inelastic and the price always declines with distance. Thisse and Vives (1988) allow for elastic consumer demand and show that prices may initially rise but could subsequently decline with distance. Firms then charge low prices to consumers in order to increase demand. High transportation costs make the demand from more distant consumers less valuable, and therefore firms charge them higher prices. For consumers located at even further distances, prices may start declining. This is because firms need to compete for distant consumers, which bids down the price levels.

[6]I consider only the following simple entry procedure: firms decide on entering sequentially in random order. Note that the order does not matter because all prospective entering firms are identical and assess their production costs based on the cross-sectional probability distribution $\{\gamma, [1-\gamma]\}$.

all consumers. Yet, if it competes with a high-cost firm, it behaves as a local monopolist for the closest consumers but competes for more distant ones.

To prevent complexity due to discreteness in the number of firms, I allow $N$ to be a continuous variable, such that $N^*$ is determined by the equilibrium condition:

$$E(\Pi(N^*)) = F. \tag{5.11}$$

This provides

**Lemma 5.2.** *The expected profits are decreasing in the number of entering firms.*

Intuitively, a higher number of firms reduces the market share of each firm, diminishing firm profits. This lemma guarantees that there exists a unique equilibrium number of entering firms $N^*$.

### 5.3.3 Social welfare

I compute social welfare as the sum of consumer and producer profits. This equals the reservation price of the products minus transportation costs, production costs, and entry costs. Social welfare per one firm is

$$W_{i,j} = \int_{-\bar{x}}^{\bar{x}} [Y - c_i - \tau x] \mathrm{d}x - F. \tag{5.12}$$

Hence, the total social welfare is

$$TW = N \times E(W_{i,j}) = N\{\gamma^2 W_{G,G} + \gamma[1-\gamma][W_{G,B} + W_{B,G}] + [1-\gamma]^2 W_{B,B}\},$$

which yields

$$TW = -\frac{\tau}{4N} + \frac{\gamma[1-\gamma]\Delta c^2 N}{2\tau} + Y - E(c) - FN. \tag{5.13}$$

## 5.4 Main Analysis

Now I analyze the effect of competition in terms of transportation costs on firms' entry decisions. I first focus on homogeneous firms. Second, I allow for firm heterogeneity. I analyze the impact of firm heterogeneity on firm profits, entry, and social welfare.

### 5.4.1 Homogeneous firms

I now analyze how the strength of competition in the market affects the market structure in the case of homogeneous firms (i.e., $c_G = c_B = c$). The expected firm profits are (rewrite (5.10))

$$E(\Pi) = \begin{cases} \frac{\tau}{2N^2} & \text{if } \tau \leq [Y-c]N, \\ \frac{2[Y-c]}{N} - \frac{\tau}{2N^2} - \frac{[Y-c]^2}{\tau} & \text{if } \tau > [Y-c]N. \end{cases} \tag{5.14}$$

Lowering production costs has the following effect on entry.

**Proposition 5.1.** *Improving cost efficiency has no impact on profits and entry at low entry costs (i.e., $F < \hat{F}$) but augments profits and entry if entry costs are high (i.e., $F \geq \hat{F}$).*

Proposition 5.1 points to two regions of competition. With low entry costs, competition is high. Improving cost efficiency has no effect on firms' profitability because the benefits are passed on to consumers. With high entry costs, competition is low. Firms behave as monopolists for closely located consumers and from them they extract bigger rents due to lower costs. Consequently, improving cost efficiency augments firm profits and entry.

I can show the following relation between transportation costs and entry.

**Proposition 5.2.** *The level of entry is a bell-shaped function of transportation costs; that is, there exists a $\bar{\tau}$ such that for $\tau < \bar{\tau}$, $\frac{\partial N}{\partial \tau} > 0$ and for $\tau \geq \bar{\tau}$, $\frac{\partial N}{\partial \tau} \leq 0$.*

The intuition for this result is the following. With low transportation costs (i.e., at $\tau < \bar{\tau}$), competition is high. Further reduction of transportation costs deprives firms of their locational advantages. That is, consumers more easily travel to a competing firm. This lowers profits and hampers entry. However, for large transportation costs an additional effect is at work. Now each firm acts as a monopolist for its closest consumers. A firm must compensate its consumers for the transportation costs that they incur. Decreasing the level of transportation costs allows firms to augment rents obtained from these consumers. The second effect prevails for a sufficiently large interval in which firms behave as a monopolists; that is, for sufficiently large transportation costs. In this case, lowering transportation costs augments profits and entry. This result replicates the analysis by Salop (1979) in the model of perfect price discrimination.

One can compute the following comparative statics with respect to $\bar{\tau}$.

**Corollary 5.1.** *The region in which transportation costs are positively related to entry is decreasing in firms' entry costs (i.e., $\frac{\partial \bar{\tau}}{\partial F} < 0$) and in production costs (i.e., $\frac{\partial \bar{\tau}}{\partial c} < 0$).*

The intuition for the first part of Corollary 5.1 is as follows. Lowering entry costs increases the level of competition, which makes transportation costs more often positively related to entry (see Proposition 5.2). That is, lowering entry costs increases the region in which transportation costs and entry are positively related.

The second part of Corollary 5.1 stems from the following effect. Lower firm production costs augment competition because firms can compete for more distant consumers. However, Proposition 5.2 showed that higher competition makes transportation costs more often positively related to entry. This shows that lower production costs augment the region in which transportation costs and entry are positively related.

### 5.4.2 Heterogeneous firms

I now allow for heterogeneity in the production costs of firms (i.e., $c_B > c_G$). First, I analyze the impact of transportation costs on entry (and the related question of the connection between firm size and demand density). Second, I focus on the interconnection between heterogeneity and competition between firms. In particular, I ask whether more heterogeneity is beneficial for firms and for social welfare.

The following proposition shows that firm heterogeneity can change the relation between competition and entry as described in Proposition 5.2.

**Proposition 5.3.** *In the case of high heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$) and for the intermediate quality of the industry (i.e., $\gamma \in (\gamma_1, \gamma_2)$), the level of entry is a $U$-shaped function of transportation costs; that is, there exists a $\tilde{\tau}$ such that for $\tau < \tilde{\tau}$, $\frac{\partial N}{\partial \tau} < 0$ and for $\tau \geq \tilde{\tau}$, $\frac{\partial N}{\partial \tau} \geq 0$.*

The intuition for Proposition 5.3 is the following. For low values of transportation costs, increasing competition through a reduction of transportation costs allows low-cost firms to take more consumers from high-cost firms. This redistribution of market share from high-cost to low-cost firms augments the efficiency of the industry. Consequently, the expected firm profits increase and entry goes up.

To understand this further, we reexamine the conditions on $\gamma$ that are necessary for the positive relation between competition (in terms of lower transportation costs) and entry. A lower bound on $\gamma$ exists because high-cost firms only lose with higher competition. They lose both their customers and rents. However, an upper bound on $\gamma$ exists as well. Low-cost firms gain with competition only if they compete with high-cost firms. Hence, there must be sufficiently many high-cost firms in the industry.

For high transportation costs, the shift of the market share is small and higher competition only lowers rents without decisively improving the efficiency. Consequently, competition lowers profits and hampers entry. In total, the level of entry is a $U$-shaped function of transportation costs.

Interestingly, involving heterogeneity in the industry may reverse the result from homogeneous industry (compare Proposition 5.3 with Proposition 5.2). In particular, in homogeneous industry lowering transportation costs results in higher competition between firms, lower expected firm profits, and lower entry. However, for high transportation costs (i.e., for $\tau > \bar{\tau}$), a perverse equilibrium exists, in which lowering transportation costs actually increases firm (almost monopolistic) profits and leads to more entry. Proposition 5.3 shows that the standard result of Salop (1979) again reverses in heterogeneous industries. In particular, lowering transportation costs may again increase expected firm profit and lead to more entry. This, now happens for low transportation costs (i.e., for $\tau < \tilde{\tau}$).

Now I analyze how increasing heterogeneity affects firm profits, their entry decisions, and social welfare. I consider the change in the heterogeneity between firms as exogeneously driven.[7] The heterogeneity in production costs extends Proposition 5.1 to the following result.

[7] In Section 5.5.2, I apply the results to the banking industry, in which heterogeneity plays an important role because of stability considerations and may be affected in part exogenously by the bank regulator.

**Lemma 5.3.** *Increasing the heterogeneity $\Delta c$ at constant average production costs $E(c)$ augments profits at low entry costs (at low $F$) but does not change profits when entry costs are high (at high $F$). The positive effect is the greatest for industries of intermediate quality (i.e., $\gamma = \frac{1}{2}$).*

At low entry costs $F$, firms fiercely compete for the consumers. This transfers almost all the rents from firms to consumers. Firms can only keep the rents if they have a cost advantage over competitors. High heterogeneity means that low-cost firms can keep higher rents when in competition with high-cost firms. This raises expected profits. The positive effect is higher if low-cost firms often compete with high-cost firms, which is the case for intermediate quality of the industry (i.e., $\gamma = \frac{1}{2}$).

However, with high entry costs, competition is low and firms largely behave as local monopolists. In this case, changing the differences in costs does not matter as long as the average reservation price $E(c)$ stays constant.

**Lemma 5.4.** *Increasing the average production cost $E(c)$ has no effect on entry with low entry costs (at low $F$) but results in lower entry when entry costs are high (at high $F$).*

At high competition (at low $F$), firms earn very small profits. Profits are mostly transfered to consumers. Increasing production costs does not affect firm profits, but results in lower consumer profits. Hence, higher production costs do not affect firm profits nor the level of entry. If competition is low, however, firms take most of the profits. Increasing production costs now hampers the profits of the firms and leads to lower entry.

Combining Lemma 5.3 with Lemma 5.4, I can derive the following result.

**Proposition 5.4.** *If increasing heterogeneity $\Delta c$ comes with higher average production costs (higher $E(c)$), entry is augmented at low entry costs $F < \hat{F}$; however, entry is hampered when entry costs are high (i.e., $F \geq \hat{F}$). A parameter $\hat{F}$ is increasing in $\gamma$ (i.e., $\frac{\partial \hat{F}}{\partial \gamma} > 0$).*

At low entry costs, competition is fierce. This makes higher heterogeneity valuable because it results in a larger shift of the market share from high-cost to low-cost firms. The efficiency of the industry is increased. Higher production costs, on the other hand, have no effect on firm profits. To see this, observe that with low entry costs competition is fierce and firms mostly pass the profits to consumers. Therefore increasing the average production cost mostly hampers consumers and does not affect firm profits. Summarizing, firm profits and entry increase.

With high entry costs, competition is low and higher heterogeneity has a limited effect on firm profits. That is, low competition confines the positive role of heterogeneity essentially by limiting a shift of the market share from high-cost to low-cost firms. Furthermore, with low competition, firms keep most of their profits. Hence, increasing production costs negatively affects firm profits. Hence, firm profits and entry decrease.

Increasing the quality of the industry $\gamma$ augments the proportion of low-cost firms in the industry. Low-cost firms are fiercer competitors than high-cost firms if everything else is equal. That is, low-cost firms can reach more distant consumers, which makes transportation costs less important (see also Corollary 5.1). Higher quality $\gamma$ then increases competition. This makes higher heterogeneity more important compared to a decrease in production costs. Summarizing, increasing heterogeneity is more valuable if the industry is of high quality.

Proposition 5.4 brings policy implications for regulated industries. Costly regulatory policies that exacerbate the differences among firms in their production costs are effective in competitive and high quality industries. They augment firm profits and entry through redistribution of market share towards more efficient firms. Yet, such policies may not be valuable in non-competitive, low quality industries. If competition is low, the shift of the market share is low and the costs of such policies decrease profits and entry.

The following corollary is intuitive.

**Corollary 5.2.** *Improving the cost efficiency of low-cost firms always augments entry.*

Lower production costs of low-cost firms increase expected profits due to two reasons. First, firms may earn higher profits due to lower production costs. Second, improving the cost efficiency of low-cost firms increases the differences in production costs between low-cost and high-cost firms. This allows low-cost firms to earn extra profits when competing with high-cost firms. Consequently, expected profits and entry increase.

**Corollary 5.3.** *Improving the cost efficiency of high-cost firms hampers entry with low entry costs (i.e., $F < \hat{F}$), but augments entry when entry costs are high (i.e., $F \geq \hat{F}$). Also $\frac{\partial \hat{F}}{\partial \gamma} > 0$.*

This counterintuitive result can be explained as follows. With low entry costs, competition between firms is fierce. The benefits of a lower production cost of high-cost firms $c_B$ are passed on to consumers (see Proposition 5.1). However, lowering the production cost of high-cost firms also decreases the differences between low-cost and high-cost firms. Lower differences increase the competitiveness of the industry. That is, firms with equal production costs can extract lower rents from consumers. In addition, lower differences between firms limit the beneficial market shift from high-cost to low-cost firms. Hence, profits and entry decrease (see Lemma 5.3). This is different with high entry costs, where firms largely behave as local monopolists. Lowering the production costs of high-cost firms now results in higher firm profits. Hence, there is more entry.

Corollary 5.2 and Corollary 5.3 point to potential differences in directing policies towards low-cost vs. high-cost firms. Whereas policies that increase the productivity of low-cost firms always benefit the industry and lead to higher entry, policies that help high-cost firms may be detrimental to the industry as a whole, especially if competition is high and the industry is of high quality. Differently said, in this case policies that further hurt high-cost firms may improve firm profits and entry.

The impact of higher differences between firms on social welfare is the following.

**Proposition 5.5.** *Increasing heterogeneity $\Delta c$ always augments social welfare.*

Proposition 5.5 is driven by a welfare-positive effect of competition. That is, competition shifts the market share from high-cost to low-cost firms. Higher heterogeneity has two positive effects. First, the shift of the market share is higher. Second, the benefits of the shift of the market share are higher. That is, a shift of one consumer from a high-cost firm to a low-cost firm is more beneficial if the differences in costs are high. Both effects increase social welfare.

Summarizing, this section first connects entry in the industry with the level of competition in the industry; namely, with the level of transportation costs. I show that for sufficiently heterogeneous firms, lowering transportation costs leads to a beneficial shift of the market share that increases firm profits and entry. Second, I show that increasing the heterogeneity of the industry augments firm profits, entry in the industry, and social welfare.

## 5.5 Extensions and Applications

In this section I first extend the analysis and connect it with the literature on optimal firm size. Second, I apply the analysis to illuminate some pertinent issues in bank regulation. I start with deposit insurance and capital regulation and continue with the role of market transparency and bail out policies.

### 5.5.1 Extension: Firm size and demand density

This analysis can also be interpreted to predict the relation between demand density and the size of firms. I show that in a sufficiently competitive and heterogeneous industry firms in denser markets may be smaller. Higher demand density affects firm size through two different channels. First, higher density creates larger demand. Larger demand induces more entry which leads to fiercer competition and lower rents. To compensate for lower rents, firms must be larger (see Syverson (2004)). However, if competition and heterogeneity of industry is high, the second channel might prevail. In particular, higher density increases competition, which allows for a greater shift of the market share from high-cost to low-cost firms and enhances the cost efficiency of the industry. The gains in cost efficiency boost profits and entry which leads to smaller firms.

The model is now expanded such that a total mass of consumers on the unit circle is given by $D$. For homogeneous firms, I can show the following proposition.

**Corollary 5.4.** *Average firm size is a $U$-shaped function of market density; that is, there exists $\bar{D}$ such that for $D < \bar{D}$, $\frac{\partial[D/N]}{\partial D} < 0$ and for $D \geq \bar{D}$, $\frac{\partial[D/N]}{\partial D} \geq 0$.*

The intuition is the following. Increasing demand density augments entry due to two effects. First, higher demand density increases demand for products, which directly augments profits and entry. An indirect effect is at work as well. That is, a higher number of entering firms lowers the distances between firms and consumers. For low levels of demand density, this actually improves profits because firms that largely behave as monopolists spend less to compensate consumers for their transportation costs. Thus, the indirect effect creates additional entry. If the indirect effect is sufficiently strong (which happens for low demand density; that is, for $D < \bar{D}$), entry in the market is so high that the number of firms grows faster than the market itself and the firm size shrinks.

At high demand density, a direct effect is still present. That is, even higher density improves firm profits which results in more entry. However, an indirect effect works in the opposite direction. In particular, a higher number of entering firms strengthens competition. Firms more fiercely compete for consumers, which leads to lower profits and entry. Consequently, higher density increases firm size.

In a similar spirit, the impact of demand density on firm size is different in a heterogeneous industry (compare with Corollary 5.4).

**Corollary 5.5.** *In the case of heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$), and for the intermediate quality of the industry (e.i., $\gamma \in (\gamma_1, \gamma_2)$), the average firm size is a bell-shaped function of demand density; that is, there exists $\bar{D}$ such that for $D < \bar{D}$, $\frac{\partial [D/N]}{\partial D} > 0$ and for $D \geq \bar{D}$, $\frac{\partial [D/N]}{\partial D} \leq 0$.*

The reason for the bell-shaped function is the following. For high demand density the relationship between demand density and firm size may be negative. This is because increasing density brings new firms into the industry, which intensifies competition. In a sufficiently heterogeneous industry, intensifying competition augments firm profits. In particular, high competition shifts the market share from high-cost to low-cost firms, which leads to more efficient industry, bigger rents, and additional entry. In total, the number of firms goes up more than the market size increases. Consequently, firms become smaller.

For low demand density, the relation between demand density and firm size is positive. This is because higher density attracts additional entry, which augments competition. However, the positive effect of competition – namely, the redistribution of the market share from high-cost to low-cost firms – is limited. Competition then aggravates firm profits, which hampers entry. In sum, an increase in demand density makes firms larger.

In an empirical analysis of the cement industry, Syverson (2004) confirms a positive relation between demand density and firm size. Syverson (2004) may account for the relation between firm size and demand density that I predict for relatively low levels of demand density. Arguably, the cement industry is geographically rather segmented. Corollary 5.5 predicts that this relation may reverse for some less segmented industry.

Now I turn to the applications of the results.

### 5.5.2 Application: Bank regulation

Heterogeneity in banking plays an important role due to its stability implications. In particular, the bank regulator may contain the heterogeneity (in the qualities) of banks to limit potentially devastating bank failures.[8] The regulator may use several regulatory tools to

[8]See Bhattacharya, Boot, and Thakor (1998) for a further rationale for bank regulation.

impact the stability of the banking system. However, such policies may also exogenously affect heterogeneity and the level of competition in the banking industry. This makes this analysis especially applicable to bank regulation. The analysis highlights the side effects of stability-oriented bank regulation that come from its impact on heterogeneity in the banking industry.[9]

Evidence from branching deregulation of the U.S. banking system identifies how intra-industry competition measured by the transportation cost affects entry in banking. Before 1970, branching in the U.S., was effectively prohibited. Beginning in the late 1970s, the U.S. gradually removed the restrictions to within-state and inter-state branching. This has substantially lowered bank-borrower distances and lowered transportation costs. Proposition 5.3 predicts that lowering the transportation costs should increase entry if the shift in the market share from low- to high-quality banks is sufficiently high (i.e., the industry is sufficiently heterogeneous and competitive). Stiroh and Strahan (2003) show that deregulation led to a substantial reallocation of market share toward better banks. Aligned with Proposition 5.3, Berger et al. (2004) confirm that entry was higher in states with newly liberalized branching rules where shifts in market share became feasible.

Proposition 5.3 also identifies a potential rationale for deregulation. Banks on their own might pressure the regulator to lower barriers to branching and, in doing so, induce more interbank competition. This applies to better quality banks that as such would gain market share. That is, such deregulation would allow for a shift in the market share toward good banks. Even with higher competitive pressure, this could elevate the better bank profitability. Kroszner and Strahan (1999) provide some evidence to this point. They show that the private-interest and the strength of winners (large banks) versus losers (small banks) affected interstate branching deregulation through influence on congressional voting.

Deposit insurance is one of the most widely used forms of bank regulation (see Demirgüç-Kunt, Karacaovali, and Laeven (2005)). Deposit insurance guarantees depositors that their claims will be repaid even if their bank fails. This may contain bank runs. However, insured depositors disregard bank risk knowing that the deposit insurer carries potential losses. Hence, depositors are willing to invest in risky and in safe banks at exactly the same interest rates. This shows that deposit insurance helps risky banks obtain a cheap source of funds: insured deposits. That is, deposit insurance puts banks on an equal footing.

In terms of this analysis, deposit insurance lowers the differences in funding costs between banks. Proposition 5.5 then presents a rather negative view on deposit insurance. In particular, deposit insurance limits an efficient redistribution of market share from high-cost (high-risk) to low-cost (low-risk) banks. This may negatively affect social welfare. Further more, Proposition 5.4 predicts that more generous deposit insurance may lead to less entry in banking, especially if competition is high and if banks are on average safe.

This shows that regulators of highly competitive and high quality banking systems should

[9] This analysis does not serve to criticize bank regulation as such. The importance of bank regulation in containing costly bank failures is unquestionable. I merely point to some of its potentially unexpected effects.

carefully consider the effect of deposit insurance. Correct implementation seems to be crucial. Although theoretical studies (see Chan, Greenbaum, and Thakor (1992)) show that a deposit insurance premium can hardly be adjusted to a bank's risk, some regulators are nevertheless moving in this direction.[10] Alternatively, the coverage of deposit insurance may be limited (see Chen (1999) and Gropp and Vesala (2004)).

Proposition 5.4 could also predict that the banking industry as a whole would lobby for limits on deposit insurance in highly competitive, high-quality banking systems. Lower deposit insurance would induce a redistribution of market share to better banks. However, in less competitive banking markets, meaning those with little competition between existing banks, the redistribution effect is limited and banks may prefer to have more generous deposit insurance.

In Chapter 3 I show that capital regulation limits the distortion of deposit insurance. In particular, capital regulation forces banks to raise capital and therefore rely less on insured deposits. Hence, risky banks benefit less due to deposit insurance. Capital regulation limits the distortions that flat-rate deposit insurance introduces to banking. Implicitly, such deposit insurance benefits lower quality banks the most, and makes them fiercer competitors than they otherwise would have been. Capital requirements are an effective regulatory tool that mitigates this distortion, and in doing so increases the value of entry.

In Chapter 3 I show that costly capital regulation may lead to higher bank profits and more entry if competition is sufficiently high. This analysis rederives this result in the framework of a spatial model. In terms of this model, increasing capital requirements augments the average cost of financing for banks ($E(c)$ increases). However, higher capital requirements at the same time increase the cost differences between banks (an increase in $\Delta c$) by limiting the effect of deposit insurance. Proposition 5.4 then shows that higher costly capital requirements may result in more entry if competition is sufficiently high; that is, if transportation costs for bank clients are sufficiently low.

This analysis also offers some implications for market-based regulation. Bank supervisors increasingly try to learn from market information when it comes to analyzing bank stability (KMV-models, Basel II, etc.). Proposition 5.4 shows that market disclosure and market pressure policies, while highlight the differences between banks, are valuable for banks only in highly competitive banking industries. That is, market-based regulation can better discriminate between low-risk and high-risk (low-cost and high-cost) banks. This increases the level of heterogeneity in the industry, leading to a shift of the market share from high-risk to low risk banks. If competition is sufficiently high, bank profits and entry are augmented. That is, the banking industry as a whole may benefit from market-based regulation. However, if the level of competition is low, the shift of the market share is limited. Hence the additional costs of market based regulation may overcome its benefits.

[10] In the U.S., the Federal Deposit Insurance Corporation (FDIC) is designing deposit insurance to reflect the risks of the financial institution more closely. The FDIC computes the deposit insurance premium using capital levels, supervisory ratings, and financial ratios and issuer ratings. See FDIC Statistics on Banking, A Statistical Profile of the United States Banking Industry, March 2006.

Corollary 5.3 applied to regulatory bailout policies points to the potentially negative effect of bailout policies even if they come at no additional cost to the banking industry. Generous bailout policies essentially help risky (and high-cost) banks obtain cheaper financing. Stable banks benefit from bailout policies substantially less. This shows that bailout policies decrease the differences between banks. If competition is high, this hampers redistribution of market share to more stable (and potentially more efficient) banks, lowering profits and entry. That is, the banking industry as a whole may be hampered by too lax bailout policies. However, if competition is low, the redistribution of the market share is limited. In this case, costless bailout policies positively affect bank stability and increase bank profits and entry.

## 5.6 Conclusions

How does higher competition in terms of lower transportation costs affect firm entry in the industry? I counter the general intuition that higher inter-industry competition results in lower profits and less entry. I show that intensifying inter-industry competition could encourage entry if firms are heterogeneous. This result comes from the efficiency gain that competition creates. In particular, higher competition induces a redistribution of consumers from high-cost to low-cost firms, augmenting firm profits and entry. I show that entry is a bell-shaped function of transportation costs if firms are sufficiently heterogeneous. Yet, in a homogeneous industry entry is a $U$-shaped function of transportation costs.

I also point to the effect of differences in production costs between firms at the level of entry. I show that costly policies that increase differences in production costs may augment entry. In particular, the shift of the market share from high-cost to low-cost firms increases the efficiency of the industry. This augments expected firm profits and entry. Interestingly, enhancing cost differences by making high-cost firms even worse may augment average firm profits and entry if the level of competition is sufficiently high.

The banking industry seems to be ideally suited for application of these results. The banking industry is highly heterogeneous and the level of competition varies substantially across countries. In addition, regulatory policies aiming at bank stability exogenously affect the level of heterogeneity of banks. I show that deposit insurance may hamper entry while capital regulation limits its negative impact on entry. Even more, higher capital requirements may lead to more entry. Surprisingly, bailout policies may hamper bank profits and entry even though they come at no additional cost for banks. As a general conclusion, the regulator should be cautious when implementing the stability-oriented regulation, especially if the banking system is rather competitive and of high quality. In this case, the distortionary effects of regulatory policies may prevail and may be harmful for the banking industry.

This analysis presents a basic attempt to study how firms' entry decisions are affected by heterogeneity and competition in a spatial model. Both heterogeneity and competition in the industry (given by the level of transportation costs) were taken as exogenously given. In Chapter 3 I make firm heterogeneity endogenous. In particular, firms (i.e., banks) could invest and improve their quality. One could also make transportation costs endogenous. For example, firms could invest and lower transportation costs for their consumers. This extension still awaits exploration.

## 5.7 Appendix

*Proof of Lemma 5.1*

I distinguish three cases. First, for $\frac{\tau}{N} \leq Y - c_B$ the expected profit of an entering bank is

$$E(\Pi) = \gamma^2 \frac{\tau}{4N^2} + \gamma[1-\gamma][\frac{\{\tau + N[c_B - c_G]\}^2}{4\tau N^2} + \frac{\{\tau + N[c_G - c_B]\}^2}{4\tau N^2}] + [1-\gamma]^2 \frac{\tau}{4N^2}. \quad (5.15)$$

Simplify (5.15) to obtain the first line in (5.10).

Second, for $\frac{\tau}{N} > Y - c_G$, one obtains

$$\begin{aligned} E(\Pi) &= \gamma^2[-\frac{\tau}{4N^2} - \frac{[Y-c_G]^2}{\tau} + \frac{Y-c_G}{N}] + \gamma[1-\gamma]\{\frac{2Y - c_G - c_B}{2N} - \frac{\tau}{4N^2} \\ &\quad + \frac{[c_B - c_G]^2 - 2[Y-c_B]^2}{4\tau} + \frac{2Y - c_B - c_G}{2N} - \frac{\tau}{4N^2} + \frac{[c_G - c_B]^2 - 2[Y-c_G]^2}{4\tau}\} \\ &\quad + [1-\gamma]^2[-\frac{\tau}{4N^2} - \frac{[Y-c_B]^2}{\tau} + \frac{Y-c_B}{N}]. \end{aligned} \quad (5.16)$$

Simplify (5.16) to obtain the second line in (5.10).

Third, for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one computes

$$\begin{aligned} \Pi_{G,B} + \Pi_{B,G} &= \frac{[Y-c_B] + [Y-c_G]}{2N} - \frac{\tau}{4N^2} + \frac{[[Y-c_G]-[Y-c_B]]^2 - 2[Y-c_G]^2}{4\tau} + \\ &\quad \frac{\{\tau + N[[Y-c_G]-[Y-c_B]]\}^2}{4\tau N^2}. \end{aligned} \quad (5.17)$$

Simplify (5.17) to obtain the third line in (5.10). ■

*Proof of Lemma 5.2*

Observe from (5.10) that for $\frac{\tau}{N} \leq Y - c_B$ one has $\frac{\partial E(\Pi)}{\partial N} < 0$. For $\frac{\tau}{N} > Y - c_G$, one computes

$$\frac{\partial E(\Pi)}{\partial N} = -\frac{\gamma[Y-c_G] + [1-\gamma][Y-c_B]}{N^2} + \frac{\tau}{2N^3}.$$

Note that Assumption (5.2) guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

For $Y - c_B \leq \frac{\tau}{N} \leq Y - c_G$ one computes

$$\frac{\partial E(\Pi)}{\partial N} = -\frac{\gamma\tau}{2N^3} - \frac{1-\gamma}{N^2}[\frac{[Y-c_B]}{N} - \frac{\tau}{2}].$$

Note that Assumption (5.2) guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$. ■

*Proof of Proposition 5.1*

For $\frac{\tau}{N} \leq Y - c$ it is known that in equilibrium $\frac{\tau}{2N^2} = F$. That is, as long as $F < \hat{F}$, where $\hat{F} = \frac{[Y-c]^2}{2\tau}$, one has $\frac{\tau}{N} \leq Y - c$. In this case $\Pi(N) = \frac{\tau}{2N^2}$. Note that $\frac{\partial \Pi(N)}{\partial c} = 0$.

For $F \geq \hat{F}$, one has $\frac{\tau}{N} > Y - c$ and $\frac{\partial \Pi(N)}{\partial c} = -\frac{2}{N} + 2\frac{[Y-c]}{\tau}$, which is negative due to Assumption 5.2. ■

*Proof of Proposition 5.2*

Differentiate (5.14) with respect to $\tau$ to obtain

$$\frac{\partial \Pi(N)}{\partial \tau} = -\frac{1}{4N^2} + \frac{[Y-c]^2}{2\tau^2} \text{ for } \tau > [Y-c]N, \tag{5.18}$$

$$\frac{\partial \Pi(N)}{\partial \tau} = \frac{1}{4N^2} \text{ for } \tau \leq [Y-c]N. \tag{5.19}$$

Note that (5.19) is always positive.

Note that (5.18) is positive for $\frac{\tau}{N} \leq \sqrt{2}[Y-c]$ but is negative for $\frac{\tau}{N} > \sqrt{2}[Y-c]$. In equilibrium it is known that

$$\frac{2[Y-c]}{N} - \frac{\tau}{2N^2} - \frac{[Y-c]^2}{\tau} = F. \tag{5.20}$$

Now compute the entry costs $\bar{F}$ for which $\frac{\partial \Pi(N)}{\partial \tau} = 0$. Use $\tau = \sqrt{2}N[Y-c]$ in (5.20) to obtain

$$\bar{F} = \frac{[Y-c]^2}{\tau}[2\sqrt{2}-2]. \tag{5.21}$$

Note that Lemma 5.2 shows that the equilibrium $N$ is decreasing in $F$. Thus, for $F < \bar{F}$ one has $\frac{\partial \Pi(N)}{\partial \tau} > 0$ and for $F \geq \bar{F}$ one has $\frac{\partial \Pi(N)}{\partial \tau} < 0$. Solve (5.21) for $\tau$ to obtain $\bar{\tau}$. Thus, $\frac{\partial \Pi(N)}{\partial \tau} > 0$ for $\tau < \bar{\tau}$ and $\frac{\partial \Pi(N)}{\partial \tau} < 0$ for $\tau \geq \bar{\tau}$. ■

*Proof of Proposition 5.3*

The proof consists of two parts. First, I show that the equilibrium lies in the region $Y - c_B < \frac{\tau}{N} < Y - c_G$. Observe that for high heterogeneity $\Delta c > \sqrt{2}[Y - c_B]$ a milder condition also holds (i.e., $\Delta c > Y - c_B$). Combine this with Assumption 5.1 to see that the region $\tau \leq [Y - c_B]N$ does not exist. More specifically, combine high heterogeneity with $\frac{\tau}{N} \leq Y - c_B$ to obtain $c_B - c_G > Y - c_B > \frac{\tau}{N}$. However, this contradicts with Assumption 5.1. Note that for high heterogeneity ($Y - c_G > 2[Y - c_B]$) there is no equilibrium for $\frac{\tau}{N} > Y - c_G$. That is, rewrite $\Delta c > Y - c_B$ to obtain $Y - c_G > 2[Y - c_B]$, which means that $\frac{\tau}{N} > Y - c_G > 2[Y - c_B]$, but this contradicts with Assumption 5.2.

Second, I show that the proof holds in the region $Y - c_B < \frac{\tau}{N} < Y - c_G$. I introduce $d = \frac{1}{N[Y-c_B]}$ and $f = \frac{F}{[Y-c_B]^2}$ and $g = \frac{\Delta c^2}{[Y-c_B]^2}$. Note that $1 < \tau d < 2$. That is, Assumption 5.2 guarantees that $\tau d < 2$ and the condition $Y - c_B < \frac{\tau}{N}$ guarantees that $1 < \tau d$. Observe also that, for high heterogeneity (i.e., $\Delta c > \sqrt{2}[Y - c_B]$), one has $g > 2$.

In what follows I prove that there exists $\tilde{F} = \tilde{f}[Y - c_B]^2$ such that for $f < \tilde{f}$ one has $\frac{\partial d}{\partial \tau} > 0$ and for $f \geq \bar{f}$ one has $\frac{\partial d}{\partial \tau} < 0$ as long as $g > [\sqrt{2}+1]^2$ and $\gamma \in (\gamma_1, \gamma_2)$, where

$$\gamma_1 = 1 - \frac{1}{g}, \quad \gamma_2 = \frac{1}{2} + \frac{\sqrt{1 - 2/g}}{2}. \tag{5.22}$$

Rewrite (5.10) for the region $Y - c_B < \frac{\tau}{N} < Y - c_G$ to obtain the equilibrium condition

$$\gamma\frac{\tau d^2}{2} + [1-\gamma][d - \frac{\tau d^2}{2} + \frac{\gamma g - 1}{\tau}] = f. \tag{5.23}$$

Differentiate (5.23) w.r.t. $\tau$ to obtain

$$\frac{\partial d}{\partial \tau} = \frac{\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]}{\tau^2\{\gamma\tau d + [1-\gamma][2-\tau d]\}}. \tag{5.24}$$

Note that $\gamma\tau d + [1-\gamma][2-\tau d] > 0$ because $\tau d < 2$. Observe that $d$ is increasing in $f$. In order to have $\frac{\partial d}{\partial \tau} > 0$ for low $f$ and $\frac{\partial d}{\partial \tau} < 0$ for high $f$, one needs $\frac{\partial d}{\partial \tau}$ to be a negative function of $d$. One needs $\gamma > \frac{1}{2}$, but this is true because $\gamma > \gamma_1$.

Observe that $\frac{\partial d}{\partial \tau} > 0$ for the lowest $\tau d$ (i.e., $\tau d = 1$). That is,

$$\left[\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]\right]_{d\tau=1} = -\frac{1}{2} + \gamma[1-\gamma]g, \tag{5.25}$$

which is positive for at least $\gamma \in (\frac{1}{2} - \frac{\sqrt{1-2/g}}{2}, \frac{1}{2} + \frac{\sqrt{1-2/g}}{2})$, and therefore also for $\gamma \in (\gamma_1, \gamma_2)$, because $g \geq 2$.

Now observe that $\frac{\partial d}{\partial \tau} \leq 0$ for the highest $\tau d$ (i.e., $\tau d = 2$). Compute

$$\left[\frac{-\gamma\tau^2 d^2}{2} + [1-\gamma][\frac{\tau^2 d^2}{2} + \gamma g - 1]\right]_{d\tau=2} = \gamma\{-1 + [1-\gamma]g\}, \tag{5.26}$$

which is negative if $\gamma > \gamma_1$. Thus, there exists $\tilde{f}$ corresponding to a certain $\tau\tilde{d}$ such that for $f < \tilde{f}$ one has $d < \tilde{d}$ and $\frac{\partial d}{\partial \tau} > 0$ and for $f \geq \tilde{f}$ one has $d \geq \tilde{d}$ and $\frac{\partial d}{\partial \tau} \leq 0$. ∎

*Proof of Lemma 5.3*

Insert $c_G = E(c) - [1-\gamma]\Delta c$ and $c_B = E(c) + \gamma\Delta c$ into (5.10). For $Y - c_B < \frac{\tau}{N} \leq Y - c_G$, one obtains

$$E(\Pi) = \gamma\frac{\tau}{2N^2} + [1-\gamma]\{2\frac{Y - E(c) - \gamma\Delta c}{N} - \frac{\tau}{2N^2} + \frac{\gamma\Delta c^2 - [Y - E(c) - \gamma\Delta c]^2}{\tau}\} - F. \tag{5.27}$$

Now differentiate profits in (5.10) (use also (5.27)) with respect to $\Delta c$ to obtain

$$\left.\frac{\partial E(\Pi)}{\partial \Delta c}\right|_{E(c)=const.} = \begin{cases} \frac{2\gamma[1-\gamma]\Delta c}{\tau} & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ 0 & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2\gamma[1-\gamma][\frac{[Y-c_G]}{\tau} - \frac{1}{N}] & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \tag{5.28}$$

Note that $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} > 0$ for $F$ sufficiently low such that $N > \frac{\tau}{[Y-c_G]}$. For sufficiently high $F$ one has $N < \frac{\tau}{[Y-c_G]}$ in which case $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} = 0$. Note also that for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one has $\frac{\partial E(\Pi)}{\partial \Delta c} > 0$. Observe also that (5.28) is the greatest for $\gamma = \frac{1}{2}$. ∎

*Proof of Lemma 5.4*

Differentiate (5.10) with respect to $E(c)$ keeping $\Delta c$ constant to obtain

$$\left.\frac{\partial E(\Pi)}{\partial E(c)}\right|_{\Delta c=const.} = \begin{cases} 0 & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ -\frac{2}{N} + \frac{2[Y-E(c)]}{\tau} & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2[1-\gamma][-\frac{1}{N} + \frac{[Y-c_B]}{\tau}] & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \tag{5.29}$$

Note that $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} = 0$ for $F$ sufficiently low such that $N > \frac{\tau}{[Y-c_B]}$. For sufficiently high $F$, we have $N < \frac{\tau}{[Y-c_B]}$ in which case $\frac{\partial E(\Pi)}{\partial \Delta c}|_{E(c)=const.} < 0$. To see this, note that $\frac{2}{N} - \frac{2[Y-E(c)]}{\tau} < 0$ because $\frac{\tau}{N} > Y - c_G$. Also observe $2[1-\gamma][\frac{1}{N} - \frac{[Y-c_B]}{\tau}] < 0$ because $Y - c_B < \frac{\tau}{N}$.

Note that $2[1-\gamma][\frac{1}{N} - \frac{Y-c_B}{\tau}]$ is maximized for $\gamma = 0$. ■

*Proof of Proposition 5.4*

We check for the sign of $DF$, where we define

$$DF = \left.\frac{\partial E(\Pi)}{\partial \Delta c}\right|_{E(c)=const.} + \alpha \left.\frac{\partial E(\Pi)}{\partial E(c)}\right|_{\Delta c=const.}. \tag{5.30}$$

Use (5.10) to compute (5.30) for $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ to obtain

$$DF = 2\gamma[1-\gamma][\frac{Y-c_G}{\tau} - \frac{\gamma}{N}] - 2\alpha[1-\gamma][\frac{1}{N} - \frac{Y-c_B}{\tau}]. \tag{5.31}$$

Use (5.10), (5.28) and (5.31) to obtain

$$DF = \begin{cases} \frac{2\gamma[1-\gamma]\Delta c}{\tau} & \text{if } \frac{\tau}{N} \leq Y - c_B, \\ -\alpha[\frac{2}{N} + 2\frac{Y-E(c)}{\tau}] & \text{if } \frac{\tau}{N} > Y - c_G, \\ 2[1-\gamma]\{\gamma\frac{[Y-c_G]}{\tau} - \alpha[\frac{1}{N} - \frac{[Y-c_B]}{\tau}]\} & \text{if } Y - c_B < \frac{\tau}{N} \leq Y - c_G. \end{cases} \tag{5.32}$$

Note that for sufficiently low $F$ this yields $\frac{\tau}{N} \leq Y - c_B$ one has (5.30) always positive. For medium $F$ such that $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ there exists

$$\bar{\gamma} = \frac{\alpha\{\frac{\tau}{N} - [Y - c_B]\}}{[Y - c_G]},$$

such that for $\gamma > \bar{\gamma}$ the value in (5.30) is positive and for $\gamma \leq \bar{\gamma}$ the value in (5.30) is negative. For very high $F$ this yields$\frac{\tau}{N} > Y - c_G$, the value in (5.30) is negative for all $\gamma$. ■

*Proof of Corollary 5.2*

In the equilibrium $E(\Pi(N^*)) = F$. Differentiate it with respect to $N^*$ to obtain

$$\frac{\partial E(\Pi)}{\partial c_G}\frac{\partial c_G}{\partial N} + \frac{\partial E(\Pi)}{\partial N} = 0.$$

Compute

$$\frac{\partial N}{\partial c_G} = -\frac{\frac{\partial E(\Pi)}{\partial c_G}}{\frac{\partial E(\Pi)}{\partial N}}.$$

Note that Lemma 5.2 guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

Now observe that for $\tau > [Y - c_G]N$ one has

$$\frac{\partial E(\Pi)}{\partial c_G} = -\gamma[\frac{1}{N} - \frac{[Y - E(c)]}{\tau}] < 0.$$

For $[Y - c_G]N > \tau > [Y - c_B]N$ one has

$$\frac{\partial E(\Pi)}{\partial c_G} = -\gamma[1-\gamma]\frac{c_B - c_G}{\tau} < 0.$$

For $\tau < [Y - c_B]N$ one also has $\frac{\partial E(\Pi)}{\partial c_G} < 0$. Thus, $\frac{\partial N}{\partial c_G}$ is always negative. ■

*Proof of Corollary 5.3*

In the equilibrium $E(\Pi(N^*)) = F$. Differentiate it with respect to $N^*$ to obtain

$$\frac{\partial E(\Pi)}{\partial c_B}\frac{\partial c_B}{\partial N} + \frac{\partial E(\Pi)}{\partial N} = 0.$$

Compute

$$\frac{\partial N}{\partial c_B} = -\frac{\frac{\partial E(\Pi)}{\partial c_B}}{\frac{\partial E(\Pi)}{\partial N}}.$$

Note that Lemma 5.2 guarantees that $\frac{\partial E(\Pi)}{\partial N} < 0$.

Now observe that for $\frac{\tau}{N} \leq Y - c_B$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = [1-\gamma][-\frac{1}{N} + \frac{Y - E(c)}{\tau}],$$

which is positive.

Note that for $\frac{\tau}{N} > Y - c_G$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = \frac{\partial E(\Pi)}{\partial E(c)}\frac{\partial [E(c)}{\partial c_B}.$$

Observe that $\frac{\partial E(\Pi)}{\partial E(c)} = 2[-\frac{1}{N} + \frac{[Y - E(c)]}{\tau}]$, which is negative due to $\frac{\tau}{N} > Y - c_G > Y - E(c)$.

Now observe that for $Y - c_G > \frac{\tau}{N} > Y - c_B$ one has

$$\frac{\partial E(\Pi)}{\partial c_B} = 2[1-\gamma]\{-\frac{1}{N} + \frac{\gamma[c_B - c_G] + [Y - c_B]}{\tau}\}.$$

That is, there exists an equilibrium $N$ where $\frac{\partial E(\Pi)}{\partial c_B} = 0$ and is defined by

$$\tilde{N} = \frac{\tau}{\gamma[c_B - c_G] + [Y - c_B]} = \frac{\tau}{Y - E(c)}. \tag{5.33}$$

This shows that there exists $\hat{F}$ such that for $F < \hat{F}$ one has $N > \hat{N}$ and $\frac{\partial E(\Pi)}{\partial c_B} < 0$, whereas for $F \geq \hat{F}$ one has $N \leq \hat{N}$ and $\frac{\partial E(\Pi)}{\partial c_B} \geq 0$. In particular, insert (5.33) into (5.10) to obtain

$$\hat{F} = \frac{[Y - E(c)]^2}{2\tau} + \frac{\gamma[1-\gamma]\Delta c^2}{\tau}. \tag{5.34}$$

Differentiate $\hat{F}$ to $\gamma$ to obtain

$$\frac{\partial \hat{F}}{\partial \gamma} = \frac{Y - c_G}{\tau} + \frac{\Delta c^2}{\tau}\{1 - \gamma + [1-\gamma]^2 + 2\gamma^2. \tag{5.35}$$

Note that (5.35) is positive. ■

*Proof of Proposition 5.5*

Differentiate (5.13) with respect to $\Delta c$ to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]N\Delta c}{\tau} + [\frac{\tau}{4N} - F + \frac{\gamma[1-\gamma]\Delta c^2}{2\tau}]\frac{\partial N}{\partial \Delta c} \tag{5.36}$$

For low entry costs, such that $\tau \leq [Y - c_B]N$, differentiate the equilibrium condition in (5.11) to obtain

$$\frac{\partial N}{\partial \Delta c} = \frac{2\gamma[1-\gamma]\Delta c N^3}{\tau^2}. \tag{5.37}$$

Insert (5.37) into (5.36) and use (5.11) to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]N\Delta c}{\tau}[\frac{1}{2} - \frac{\gamma[1-\gamma]\Delta c^2 N^2}{\tau^2}]. \tag{5.38}$$

Note that Assumption 5.1 guarantees that (5.38) is always positive.

For $Y - c_B < \frac{\tau}{N} \leq Y - c_G$ one can see that

$$\frac{\partial N}{\partial \Delta c} = N\frac{2\gamma[1-\gamma][\frac{[Y-c_G]}{\tau} - \frac{1}{N}]N}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y - c_B] - \frac{\tau}{N}]} < N\frac{2\gamma[1-\gamma]\frac{\Delta c}{\tau}N}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y - c_B] - \frac{\tau}{N}]}. \tag{5.39}$$

Note also that

$$N[\frac{\tau}{2N^2} - F + \frac{\gamma[1-\gamma]\Delta c^2}{2\tau}] = \frac{\tau}{4N}[3 - 4\gamma] - [Y - c_B][1-\gamma][2 - \frac{[Y - c_B]N}{\tau}] - \frac{\gamma[1-\gamma]\Delta c^2 N}{\tau}. \tag{5.40}$$

Insert (5.39) and (5.40) into (5.36) to obtain

$$\frac{\partial TW}{\partial \Delta c} > \gamma[1-\gamma]N\frac{\Delta c}{\tau}\frac{\gamma\frac{\tau}{N} - \frac{\tau}{4N} + [1-\gamma]\frac{[Y-c_B]^2 N}{\tau} - \frac{\gamma[1-\gamma]\Delta c^2 N}{\tau}}{\gamma\frac{\tau}{N} + [1-\gamma][2[Y - c_B] - \frac{\tau}{N}]}. \tag{5.41}$$

Note that Assumption 5.1 guarantees that (5.41) is always positive.

Observe that for $\frac{\tau}{N} > Y - c_G$ the equilibrium number of firms as determined with (5.11) does not change with $\Delta c$. Consequently, differentiate social welfare in (5.13) to obtain

$$\frac{\partial TW}{\partial \Delta c} = \frac{\gamma[1-\gamma]\Delta c N}{\tau},$$

which is always positive. ■