Submission Number: PET11-11-00182

Selection, heterogeneity and entry in professional markets

Claudio Piga
Loughborough University

Marco Alderighi
University Valle d'Aosta

Abstract

We analyze two different cases of entry regulation in professional markets: first, when licensing is a requirement for becoming a professional (lawyers); second, when entry and price restrictions are applied on a geographical basis (pharmacists). Both cases are investigated within a circular model of localized competition and heterogeneous players. The analysis reveals that licensing introduces a selection mechanism which is effective in preventing entry of inefficient players in markets with large ex-ante heterogeneity. Furthermore, because in the second case excessive entry is reduced as the degree of heterogeneity increases, our analysis lends support to a policy that simultaneously relaxes entry and price restrictions.

Submitted: March 03, 2011.

Selection, Heterogeneity and Entry in Professional Markets*

Marco Alderighi[a]
Università della Valle d'Aosta, Italy.
Università Bocconi, Milano, Italy.

Claudio A. Piga[b]
Loughborough University, Loughborough, UK.
RCEA, Rimini, Italy.

October 26, 2010

Abstract

We analyze two different cases of entry regulation in professional markets: first, when licensing is a requirement for becoming a professional (lawyers); second, when entry and price restrictions are applied on a geographical basis (pharmacists). Both cases are investigated within a circular model of localized competition and heterogeneous players. The analysis reveals that licensing introduces a selection mechanism which is effective in preventing entry of inefficient players in markets with large ex-ante heterogeneity. Furthermore, because in the second case excessive entry is reduced as the degree of heterogeneity increases, our analysis lends support to a policy that simultaneously relaxes entry and price restrictions.

JEL classification: L11, D61.

Keywords: Liberal professions; Cost/quality heterogeneity; Entry regulation; Localized competition.

*A previous version of this paper appeared under the title "The circular city with heterogeneous costs". The authors would like to thank Paul Dobson, Andrè de Palma, Francesco Passarelli, Joanna Poyago-Theotoky, Jacques Thisse, Chris Wilson and the participants of 2010 IIO Conference in Vancouver, of the 2010 EARIE Conference in Instanbul, of the CCRP Workshop organized by The City University, London, and all seminar participants at CORE (Belgium) and York (UK) for their suggestions and comments. The usual disclaimer applies.

[a]E-mail: m.alderighi@univda.it

[b]Corresponding Author. E-mail: c.a.g.piga@lboro.ac.uk

1 Introduction

The degree of efficiency in markets for such liberal professions as lawyers, notaries, architects, engineers and pharmacists has been the subject of extensive investigation in Europe (European Commission, 2004 and 2005).[1] The high level of regulation characterizing the European market, in the form of either State regulation or self-regulation by professional bodies, has been deemed by some as unnecessary and harmful, while by others as compelling and vital. For the latter, regulation is usually sustained by arguments associated with the so-called "public interest view" (i.e., regulation addresses market failures due to asymmetric information, and/or externalities and/or public good provision). In contrast, those supporting the "private interest approach"claim that, due to regulatory capture, many regulatory mechanisms may serve the private interests of professional bodies' members more than those of the general public (Becker, 1983; Peltzman, 1976; Philipsen, 2009; Posner, 1974; Stigler, 1971).

To understand whether some restrictive practices should be removed or not, it is therefore essential to shed more light on the functioning and on the properties of these (regulated) markets. Based on the premise that in the real world some degree of firm heterogeneity is the rule rather than the exception, in this study we deviate from the usual market failure versus regulator capture arguments to provide a different and, in some sense, complementary model for liberal professions, which allows an evaluation of the welfare properties associated with 1) the entry requirement of licensing, 2) the entry restrictions capping the maximum number of firms and 3) fixed or minimum prices.[2]

Even within the European Union, professional regulation presents significant differences among countries and professions, although there are also some important similarities worldwide. As far as entry restrictions are concerned, for all professions a university degree in the relevant field is required; for a subset of them (lawyers) the exercise of the profession is conditional on the further acquisition of a licence which is obtained by passing an examination and on spending a period of apprenticeship under the supervision of a qualified professional (Kleiner, 2000). For other professions, e.g. pharmacists, no licensing is needed; in this case, however, rules on entry based on demographic and geographic criteria often make it impossible to open a new outlet in markets where such criteria are

[1]According to the EU Directive on Recognition of Professional Qualifications (2005/36/EC), liberal professions are "those practised on the basis of relevant professional qualifications in a personal, responsible and professionally independent capacity by those providing intellectual and conceptual services in the interest of the client and the public".

[2]The other categories of potentially restrictive regulation discussed in European Commission (2004 and 2005) and in Philipsen (2009) are: advertising regulations and regulations governing business structure, fixed and recommended prices, and multi-disciplinary practices.

binding (Schaumans and Verboven, 2008).[3]

We propose a set-up that accommodates two main cases based on whether or not a licence is required for the entry in the liberal profession. When licensing is not required, entry is modelled as a two-stage game, where prospective entrants incur a set-up sunk cost before starting production and the related competitive stage. To account for licensing, we introduce a preliminary stage, where prospective professionals have to incur the licensing sunk cost before deciding whether to start their activity or not. The opportunity cost of the time spent to complete the apprenticeship period largely determines the magnitude of the licensing cost. Thus, entry in professions where licensing is required entails a three-stage process and two different sunk costs (i.e., the licensing cost as well as the production set-up cost as in the no-licensing case)

Both types of entry games (i.e., with licensing and without) are analyzed in this study by modeling the relevant market as a circular city model of localized competition with heterogenous costs' firms (Salop, 1979; Syverson, 2004; Vogel, 2008). To remain in keeping with the terminology used in the existing literature, we will use the terms "professional" and "firm" interchangeably. Moreover, we will refer to differences in the quality between competing professionals as differences in firms' costs. The link comes as, in order to perform with the same quality level, a professional with a lower intrinsic ability has to exert a higher effort, and therefore sustain higher costs than a professional with higher ability.

The two games have different informational structures, which impact on the characteristics of the ensuing equilibrium. In the two-stage game, before moving to the final stage, players have no information about their own production costs (ability), while in the licensing game, a professional learns her own cost after the licensing phase. Thus, this second model entails a selection mechanism: only those professionals with sufficiently low production costs (high ability) will decide to enter the production stage by paying the second set-up sunk cost.

We study the properties of the selection mechanism by focusing on how it responds to differences across markets in their supply-side characteristics. In particular, we investigate whether the upper threshold of the ex-post cost distribution determined in the selection stage reacts to the degree of ex-ante cost heterogeneity characterizing the potential entrants in a market. That is, we ask whether the selection mechanism is capable to weed out the least efficient professionals precisely in those markets where there is a larger risk that relatively inefficient ones may enter.

[3] Licensing and quantitative entry restrictions are not always mutually exclusive. For instance in Italy they are both used to regulate entry into the notary profession, which is generally found to be highly regulated across Europe (Philipsen, 2009).

Since the selection mechanism occurs only in the licensing game, we thus investigate a potentially beneficial effect of maintaining licensing as an entry requirement in liberal professions. Our findings indeed point out that following the licensing stage, the selection mechanism induces a truncation from above of the ex-ante cost distribution and that such truncation intensifies as the degree of ex-ante cost heterogeneity increases. That is, the maximum level of cost draw guaranteeing non-negative profits for an operative professional is inversely related with the ex-ante cost variance. Our model thus predicts that only the professional with higher ability (i.e. lower costs) will decide to become a liberal professional and those with lower ability (i.e. higher costs) will move to other jobs as, e.g., in Leland (1979): "Doctors (or potential doctors) [...] may not be willing to remain in (or enter) the market.. (p. 1329)". Thus, our analysis suggests that licensing may play a crucial role in determining high levels of efficiency across markets, and especially in those markets where the competence base of potential professionals is relatively more dispersed. Furthermore, this result provides an additional rationale for licensing, which, unlike previous contributions, is obtained in the absence of informational asymmetry between the service providers and the buyers (Kleiner, 2000; Leland, 1979; Shapiro, 1986; Philipsen, 2009).

This model can be applied more in general to describe the entry decisions in markets with localized competition. With this respect, our three-stage entry game is related with the work of Syverson (2004). Apart from some differences in the assumed informational structure of the game (in Syverson (2004) firms decide to operate and set their prices without knowing their competitors' cost draws, while firms in our analysis set prices by knowing all costs and the relative locations of all their competitors), the focus of the two papers is rather different. Syverson (2004) investigates how such demand-side characteristics as market density can determine large productivity differences across markets. While our analysis also accommodates such demand-driven effects, our main focus is on the largely unexplored combined role of a regulatory mechanism (i.e., licensing) and supply-side characteristics (i.e., cost heterogeneity).

In the two-stage game without licensing we study the inter-play between a type of regulatory restriction on entry and pricing, by focussing in particular on the insights that can be gained by extending the analysis to the case of cost heterogenous firms. Deneckere and Rothschild (1992) highlight the general property of excessive entry in circular city model of localized competition amongst homogenous cost firms. However, Spulber (1995) shows that asymmetric information acts as a stimulus to competition since it provides an incentive to incur the set-up cost of entry. By introducing heterogeneous firms who decide whether to enter without knowing their rivals' costs, we show that excessive entry reduces as the degree of cost heterogeneity increases. This is because the first-best number of firms increases with cost heterogeneity faster than the free-entry one. Therefore, the gap between the number of firms in the free-entry market equilibrium and the socially optimal number of firms shrinks as cost heterogeneity expands.[4] Our analysis also indicates that uniform pricing, which may arise from recommended and fixed prices by professional bodies, is likely to be a particularly harmful form of regulatory restriction.

The analysis of the two-stage game suggests the following policy implications. When the first-best outcome is too costly to implement, the second best entails that price setting should be free so as to reflect firms' relative costs as much as possible, while entry restrictions, for instance those based on a geographic area population, may be relaxed. Schaumans and Verboven (2008), using data on the Belgian markets for pharmacies, focus their counterfactual analysis on these two aspects. They find that a relaxation of the entry restrictions aimed at doubling the maximum number of pharmacies per capita, coupled with a non-uniform change in the regulated mark-up (i.e., the mark-up is reduced by a smaller amount for pharmacies in rural areas, which have presumably higher costs), would leave market coverage unchanged but would generate large shift in rents to consumers.

The next Section describes the model's set-up and characterizes the equilibrium prices as well as the condition for a unique Nash equilibrium in pure strategies. The three-stage entry game with a selection mechanism is developed in Section 3, which is followed by the two-stage entry game without licensing. Section 5 concludes. The Appendix contains the proofs for all the Propositions, Lemmas and Corollaries not given in the text.

2 The model

Consider a circular city of unitary length with uniform density D. There are $N \geq 2$ equidistant firms, whose location is indexed by the sequence of integers:

$$n \in \mathbb{L} = \left\{ \underline{l}\left(N\right), \underline{l}\left(N\right) + 1, \cdot\cdot, 0, 1, \cdot\cdot, \bar{l}\left(N\right) \right\},$$

where $\underline{l}(N) = -\bar{l}(N) = -\frac{N-1}{2}$ if N is odd, and $\underline{l}(N) = -\frac{N}{2} + 1 = -\bar{l}(N) + 1$ if N is even.[5] The number of firms is endogenously determined within the model.

Firms have identical unit variable costs of production, c, but offer a service characterized by a different, exogenously determined level of quality, θ_n, $n \in \mathbb{L}$, with $\theta_n \in [\theta_L, \theta_H]$.

[4] See also Gu and Wenzel (2009), for a circular city model where excessive entry is eliminated if consumers have a sufficiently elastic demand curve.

[5] For example, if $N = 12$, then firms are denoted by the sequence $\left\{ \underline{l} = -5, -4, \cdot\cdot, 0, 1, \cdot\cdot, \bar{l} = 6 \right\}$. Indexing the firms in such a way halves the weights used in the vector of equilibrium prices, see below.

Heterogeneity is therefore a direct consequence of the intrinsic abilities of each professional in the market, which we assume to be positively correlated with the quality of the service. Throughout the paper we assume that firms compete by setting their price when costs, qualities and locations are common knowledge.

The utility a consumer, who is located at distance $d \in [0, 1/N]$ from firm n, obtains from buying a service is $U_{d,n} = v - td - \widehat{p}_n + \theta_n$, where t is the unit transport cost, $\widehat{p}_n$ is the uniform price charged by firm n and the quality of firm n's service, θ_n positively affects consumers' utility. We assume that the reservation value v is sufficiently high so that each consumer buys a unit of the good and the market is fully covered.

Let $\widehat{p}_n$, $\widehat{p}_{\langle n-1\rangle}$ and $\widehat{p}_{\langle n+1\rangle}$ denote the prices charged by firm n and its two immediate neighbours.[6] The marginal consumer between n and $\langle n+1\rangle$ ($\langle n-1\rangle$) lies at a a distance S_n^R (S_n^L) from firm n, where

$$S_n^R = \frac{\widehat{p}_{\langle n+1\rangle} - \theta_{\langle n+1\rangle} - \widehat{p}_n + \theta_n}{2t} + \frac{1}{2N}; \quad S_n^L = \frac{\widehat{p}_{\langle n-1\rangle} - \theta_{\langle n-1\rangle} - \widehat{p}_n + \theta_n}{2t} + \frac{1}{2N} \tag{1}$$

Standard computations yield the set of first order conditions (Tirole, 1988, p. 283):

$$\widehat{p}_n - \theta_n = \frac{2(c - \theta_n) + \widehat{p}_{\langle n-1\rangle} - \theta_{\langle n-1\rangle} + \widehat{p}_{\langle n+1\rangle} - \theta_{\langle n+1\rangle}}{4} + \frac{1}{2}\hat{k}, \ \forall n \in \mathbb{L}. \tag{2}$$

This problem is isomorphic to one where firms' heterogeneity is modeled by allowing firms to have different unit variable costs $c_n \geq 0$ with $c_n = c - \theta_n \in [c_L = c - \theta_H, c_H = c - \theta_L]$ and $p_n = \widehat{p}_n - \theta_n$ for any $n \in \mathbb{L}$. Bearing in mind the equivalence between quality and cost heterogeneity, to save notation in the remainder of the paper we will assume that firms differ in their cost efficiency.

2.1 Price Equilibrium

Following Eaton and Lipsey (1978), we keep the requirement that competition is localized, that is, each firm competes with its two adjacent firms on the captive market represented by the closer consumers located between them.

Condition 1 *(No mill-price undercutting)* $S_n^R \geq 0$; $S_n^L \geq 0$, *for* $n \in \mathbb{L}$.

[6]To deal with the boundaries imposed in $\mathbb{L}$, denote the number of locations separating any two firms as $i \in \mathbb{L}$. The operator $\langle\rangle$ is such that $\langle n+i\rangle \in \mathbb{L}$ and:

$$\langle n+i\rangle = \begin{cases} n+i & \text{if } \underline{l} \leq n+i \leq \bar{l} \\ n+i-N & \text{if } n+i > \bar{l} \\ n+i+N & \text{if } n+i < \underline{l} \end{cases}$$

Firms do not reduce prices in such a way to grab all the market of one or both 1-step neighbors. From a technical viewpoint, Condition 1 rules out the possibility that an inefficient firm is driven out of the market when it competes against highly efficient neighboring rivals; it therefore imposes a limit on the maximum allowable amount of cost heterogeneity in the model (Alderighi and Piga, 2010).

The price equilibrium with heterogenous firms presents a number of characteristics which do not feature in the standard, homogenous firms case. Such characteristics are now briefly illustrated, since they describe the equilibrium of the last stage of the two entry games on which we focus the attention in this study.

Lemma 1 *If Condition 1 is satisfied, then system (2) has a unique solution. The market equilibrium prices are the sum of the mark-up $\hat{k}$ and of a weighted average of all the firms' costs:*

$$p_n^* = \sum\nolimits_{i=\underline{l}}^{\overline{l}} w_{\overline{l}-|i|} c_{\langle n+i \rangle} + k,\ n \in \mathbb{L}. \tag{3}$$

Despite the localized competition assumption, the market equilibrium arises from the interaction of a sequence of chain-linked, inter-locked sub-markets (Chamberlin, 1933, p. 103-104; Rothchild, 1982), where each firm's cost affects the pricing of both direct and non-direct competitors through a transmission mechanism whose properties are defined by the weights in the following Lemma.

Lemma 2 *For any $N \geq 2$ the weights w_i and k in Lemma 1 are:*

$$w_i = 4w_{i-1} - w_{i-2} \text{ for } i = 1, ..\overline{l} - 1 \tag{4}$$

$$w_1 = \left(3 - \underline{l} - \overline{l}\right) w_0 \tag{5}$$

$$4w_{\overline{l}} = 2 + 2w_{\overline{l}-1} \tag{6}$$

$$\sum\nolimits_{i=\underline{l}}^{\overline{l}} w_{\overline{l}-|i|} = 1, \tag{7}$$

$$k = \hat{k} = t/N. \tag{8}$$

First, weights decrease with distance, so that a shock in a firm's costs propagates throughout the market but has a larger impact on the price of its closer competitors. Second, because weights sum to one, (3) is a generalization of the price equilibrium in the standard case with identical cost firms. Third, weights are independent of the cost differential between any pairs of firms, implying, for instance, that a high-cost firm will assign the same weight to its own cost regardless of whether it faces a high- or low-cost direct competitor.[7]

Lemma 3 *Equations (3)-(8) represent a unique Nash price equilibrium in pure strategies when Condition 1 is satisfied, i.e., if:*

$$c_H - c_L < \rho_c(N)\,\hat{k}, \tag{9}$$

where $\hat{k} = t/N$ *and* $\rho_c(N) = \left(w_{\bar{l}} - w_0\right)^{-1}$.

Eq. (9) indicates the maximal cost heterogeneity such that there always exists an indifferent consumer located between two neighboring producers, i.e. a high cost firm facing tough competition from two low-cost neighboring rivals always has a non-negative market share.

Unlike the homogenous firms' case, the vector of market prices (3) induces a distortive allocative outcome where consumers who should patronize a low-cost firm end up buying from a higher cost firm. Indeed, for any N:

Lemma 4 *The first-best full information allocative solution can be obtained by setting:*

$$p_n^F = c_n + k_F, \tag{10}$$

where k_F *can be set freely with the only caution that the delivered price does not exceed the reservation price of consumers.*

When firms have heterogeneous costs their optimizing behavior induces them not only to base their prices on their own costs (as in the first-best situation) but also on those of the opponents. That is, for $k_F = \hat{k}$, efficient low-cost firms charge prices above the first-best solution and inefficient high-cost ones generally charge prices below it.

3 Entry with licensing

In many countries, before an individual is allowed to practice a profession, a certain set of regulatory conditions must be fulfilled (Philipsen, 2009). Examples of measures include minimum periods of education and/or professional experience, mandatory registration, establishment requirements, licensing. Acquiring a licence usually entails the fulfilment of two mandatory requirements: passing of an examination testing the necessary entry level

[7] To calculate the numerical values of the weights in (4)-(7), we use the following approximation which holds for any i when N is large (proof is available on request):$w_{\bar{l}} = y = 1/\sqrt{3}$; $w_{\bar{l}-i}/w_{\bar{l}-i-1} = x = 2+\sqrt{3}$; $w_{\bar{l}-i} = y \cdot x^{-i}$.

competencies, and an apprenticeship period during which the entrant 1) works under the supervision of an experienced professional and 2) may be required to attend compulsory training courses.[8] Entry restrictions may however vary between professions; for instance, licensing is generally compulsory in the legal and medical profession (Philipsen, 2009; Shapiro, 1986), but it is not for pharmacists (Schaumans and Verboven, 2008).

Therefore, depending on the profession, there are differences in the type of ex-ante expenditures that a potential entrant has to sink before entering the profession. To start production, all professions incur a sunk set-up cost. This could include the present discounted value of fees paid for the mandatory membership to Professional Bodies, the time spent to set up the practice, to hire collaborators, etc. Prior to this, in some professions licensing imposes an extra sunk cost that has to be borne before an individual knows whether s/he will obtain the licence. For instance, in December 2009 the UK Social Work Task Force published a report proposing that, to be allowed to practice, social workers in Great Britain should be required to qualify for a licence after a probationary year following the completion of a degree.

The licensing cost arises as a consequence of a number of aspects. While the cost to acquire a university degree is not sunk, as the degree can be used to pursue a number of alternative career options, during the apprenticeship period a prospective professional, e.g., someone who has recently gained a degree in Law, has to combine the general know-how obtained from the university studies with the acquisition of the knowledge of both job-specific skills and practical aspects of the profession. This process, which determines a professional's intrinsic ability, is highly idiosyncratic and therefore is unlikely to yield a purely deterministic outcome, in the sense that, for instance, individuals with good academic credentials may not necessarily be very apt to tackling the more practical or relational aspects of the profession. Indeed, as noted by Holmstrom (1999), productive abilities are revealed over time through the observation of performance, and therefore by the end of the apprenticeship period, prospective professionals become aware of their overall level of intrinsic ability. The magnitude of the licensing cost is therefore represented by the opportunity cost of foregoing remunerated job opportunities during the apprenticeship period. This is assumed to be the same for all the prospective professionals, since they all have an identical outside option, as it would be the case for graduates with little or no previous work experience.

Based on the foregoing discussion, we model entry in liberal profession as either a three-stage or a two-stage game, depending on whether the licensing requirement is present or

[8] See, for instance, the Licensing Process for lawyers in Upper Canada - http://rc.lsuc.on.ca/jsp/licensingprocesslawyer/index.jsp

not. The former case is developed in this Section, while the latter in the next one. The last stage of both type of games corresponds to the price competition of the previous Section.

There is an important difference between the two games. Licensing entails the acquisition of information on a firm's own level of efficiency which may induce some firms to abandon the market before the actual entry stage, after they compare their marginal cost (or intrinsic ability) with the average marginal cost that is expected to prevail in the market. In other words, licensing triggers a selection mechanism that restricts access to those firms whose cost levels fall in the lower part of the ex-ante industry distribution of costs. While this truncation of the cost distribution from above has been already discussed within a different informational set-up in Syverson (2004), a novel result in this Section highlights the effectiveness of the selection mechanism in relation to the degree of firm heterogeneity in the market. More precisely, we show that the cost threshold below which actual entry takes place decreases as the ex-ante cost heterogeneity increases.

The timing of the licensing game is as follows.

Stage 1 **- Licensing;** Out of a pool of Λ potential candidates, $M \leq \Lambda$ prospective entrants decide to fulfill the legal requirement of obtaining a licence to exercise a given profession; by investing F_L they acquire private information on their own cost but not on that of their rivals;

Stage 2 **- Selection and actual entry;** Firms decide whether to exit (i.e., they pursue an outside option) or to enter the professional market by incurring a set-up cost F_p to start production;

Stage 3 **- Full Information Pricing;** Prices are set under a full information scenario; i.e., according to the Lemmata 1 and 2.

The first stage reflects the fact that not all the individuals with a relevant degree seek to become licensed professionals; stage 2 entails that some firms abandon the market even if they have fulfilled all the necessary formal requirements. For example, in the legal profession it is not uncommon to observe individuals opting not to enter the profession even after obtaining the licence.

There is an important difference between the structure of our game and that presented in Syverson (2004) where firms set their prices based only on their cost type but not on their rivals' actual cost realizations. While this may be a realistic assumption in the short-run, our approach closely mimics a complete information, long-run equilibrium where firms have learnt to set prices from which they would not unilaterally want to deviate (which, given Lemmata 1 and 2, implies a full knowledge of all the firms' costs and locations as in Vogel (2008)).

To derive the properties of the entry game equilibrium, we assume that costs are identically and independently distributed:

Assumption 1 *Let $\tilde{c}_n \in [c_L, c_H]$ be a random variable with distribution G, mean value $E[c_n] = \bar{c}$, variance $E\left[(c_n - \bar{c})^2\right] = \sigma^2$ and covariance $E[(c_n - \bar{c})\, c_m] = 0$, $\forall n, m \in \mathbb{L}$ and $n \neq m$.*

We solve the model by backward induction. The firms' profits in stage 3 are computed using Lemmata 1 and 2 and the restriction (9) on maximal cost heterogeneity is maintained.

3.1 Selection

In the second stage, a generic firm n knows c_n, M and the prior distribution of costs G. Because costs are randomly and independently distributed, learning its own cost does not change a firm's beliefs on its rivals' costs. Nonetheless, in the selection phase firms can form their beliefs as to the cost distribution that will emerge in the price competition stage, where only a subset $N \leq M$ of firms may be involved. Based on such beliefs, each firm assumes a posterior distribution of the opponents' costs G_e, which reflects each firm's beliefs on its rivals' decision on whether to stay or exit the market. That is, the prior and posterior distribution may differ because with a large cost heterogeneity, the producers drawing a sufficiently high cost expect to gain profits that are below the fixed set-up cost F_p and therefore abandon the market before entering production.

Under the assumption that there is at least one opponent that is going to enter the market, the expected profit (gross of pre-entry fee F_L) of firm n, if it enters the market (together with $N-1 \geq 1$ other competitors) is:

$$\begin{aligned} E_{G_e}\left[\tilde{\Pi}_n | c_n, N, M\right] =& \frac{D}{2t} \cdot E_{G_e}\left[(\tilde{p}_n - c_n) \cdot \left(\tilde{p}_{\langle n-1\rangle} + \tilde{p}_{\langle n+1\rangle} - 2\tilde{p}_n + 2\hat{k}\right)\right] - F_p \\ =& \frac{D}{t} \cdot E_{G_e}\left[(\tilde{p}_n - c_n) \cdot \left(\tilde{p}_{\langle n-1\rangle} - \tilde{p}_n + \hat{k}\right)\right] - F_p, \end{aligned} \tag{11}$$

where E_{G_e} means that the expectations are taken using firm n's posterior cost distribution on the actual entrants, G_e and the superscript $\sim$ denotes stochastic variables. The second line in (11) derives from the fact that in stage 2 firm n treats $\tilde{p}_{\langle n-1\rangle}$ and $\tilde{p}_{\langle n+1\rangle}$ in an identical manner. Further, even if firm n is a monopolist, its profit must be bounded from above, since it is limited by the consumers' willingness to pay v. Therefore:[9]

$$E_{G_e}\left[\tilde{\Pi}_n|c_n,1,M\right]=D\left(v-c_n-t/2\right)-F_p<\infty. \tag{12}$$

Define $\bar{c}_e=E_{G_e}\left(c_{\langle n+i\rangle}\right)$, $\sigma_e^2=E_{G_e}\left(c_{\langle n+i\rangle}\left(c_{\langle n+i\rangle}-\bar{c}_e\right)\right)$, $\forall i\neq 0$.

Lemma 5 *Under Assumption 1,*

a) for $N=1$, the expected profit of firm n is given by (12); for $N\geq 2$:

$$E_{G_e}\left[\tilde{\Pi}_n|c_n,N,M\right]=\frac{D}{t}\left(\left(w_d\left(\bar{c}_e-c_n\right)+\hat{k}\right)^2+W_L\left(N\right)\sigma_e^2\right)-F_p, \tag{13}$$

where $w_d=\left(1-w_{\bar{l}}\right)=\left(w_{\bar{l}}-w_{\bar{l}-1}\right)$, $W_L\left(N\right)=\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\left(w_{\bar{l}-|i+1|}-w_{\bar{l}-|i|}\right)>0$, $\mathbb{L}_0\equiv\mathbb{L}\backslash\left\{0\right\}$, $\lim_{N\to\infty}W_L\left(N\right)=\bar{W}_L=\frac{2}{9}\sqrt{3}-\frac{1}{3}\simeq 0.051567$.

b) $E_{G_e}\left[\tilde{\Pi}_n|c_n,N,M\right]$ is decreasing in c_n and F_p, and increasing in $\bar{c}_e$, σ_e^2 and D.

Values of $W_L\left(N\right)$ are given in Table 1. Provided that cost differentials are sufficiently small, Lemma 5 holds for each firm and for each possible N. Thus, sequential rationality leads to the conclusion that each firm expects that the opponents deciding to enter the market are sufficiently efficient, i.e., their cost is not greater than a threshold level α. Therefore, the posterior distribution of such firms' costs corresponds to the prior distribution $G(x)$ truncated at α. That is, $G_e\left(x\right)=G_\alpha\left(x\right)=\min\left\{G\left(x\right)/G\left(\alpha\right),1\right\}$.

Conjectures on $\alpha\in\left[c_L,c_H\right]$ also affect the number of firms that participate to the pre-entry stage: $\tilde{N}=\tilde{N}\left(\alpha,M\right)$. Since cost distributions are independent, we compute the probability that there are $\eta\leq M$ (including n) firms whose costs fall under the entry threshold α:

$$\omega_\eta=\Pr\left(\tilde{N}=\eta\right)=G\left(\alpha\right)^\eta\left(1-G\left(\alpha\right)\right)^{M-\eta}.$$

It follows that if firm n decides to enter the market, provided that in the first stage there are M firms, its expected profit is:[10]

$$\sum\nolimits_{\eta=1}^{M}\omega_\eta E_{G_\alpha}\left[\tilde{\Pi}_n|c_n,\eta,M\right], \tag{14}$$

which represents the expected profit in the selection phase for given c_n, M and α. Equivalently, for M large, the central theorem guarantees that $\tilde{N}(\alpha,M)\to N(\alpha,M)=G\left(\alpha\right)M$

[9] The case of monopoly can generate problems of existence of equilibria when the market is too small for two firms but large enough for one firm. See for example Levin and Peck (2003) for a detailed analysis on entry with at least two potential entrants.

[10] When firm n decides to enter the market it has costs lower than or equal to α and expects that the other firms entering the market have costs lower than or equal to α.

since $\omega_N \to 1$; (14) becomes:

$$E_{G_\alpha}\left[\tilde{\Pi}_n|c_n, M\right] = E_{G_\alpha}\left[\tilde{\Pi}_n|c_n, G(\alpha)M, M\right] \tag{15}$$

In the subsequent analysis, we assume M large so that the expected profit can be approximated by equation (15). Note that $E_{G_\alpha}\left[\tilde{\Pi}_n|c_n, M\right]$ retains most of the properties of $E_{G_e}\left[\tilde{\Pi}_n|c_n, N, M\right]$ in Lemma 5, i.e.

Lemma 6 $E_{G_\alpha}\left[\tilde{\Pi}_n|c_n, M\right]$ *is increasing in D and decreasing in c_n, F_p and M.*

On the one hand, we are mainly interested in the comparative static analysis concerning the impact of a change in cost heterogeneity on the expected profit; on the other, this is not a straightforward task to investigate because the expected profit in (15) is not only affected by the variance of the prior cost distribution σ, but it depends on $\bar{c}_\alpha$, σ_α and $G(\alpha)$, i.e., the mean, the variance and the shape of the posterior distribution. That is, a mean-preserving change of the distribution G^σ in $G^{\acute{\sigma}}$, with $\acute{\sigma} \neq \sigma$, may produce different results depending on how the probability mass is distributed. One way to study a change in heterogeneity is to assume a stretch of the initial distribution G^σ to obtain $G^{\acute{\sigma}}$.

Assumption 2 *Let $\mathcal{G} = \{G^\sigma, \sigma \in (0, \hat{\sigma}]\}$ is a family of bounded distributions parameterized by σ, twice continuously differentiable, with differentiable density g^σ, average $\bar{c}$ and variance σ^2, such that $G^\sigma(x) = G^s\left(\frac{x-\bar{c}}{\sigma}\right)$, where G^s has mean value 0 and variance 1. In addition, let $\phi(\alpha, c_n, M, \sigma) = E_{G^\sigma_\alpha}\left[\tilde{\Pi}_n|c_n, M\right]$, and $\Phi(\alpha, M, \sigma) = \phi(\alpha, \alpha, M, \sigma)$. We require that:*

(A) $G \in \mathcal{G}$.

(B) (monotonicity) $\phi(\alpha, c_n, M, \sigma)$ *is decreasing in α and σ.*

(C) (single crossing) $\frac{\phi_\alpha(\alpha,\alpha,M,\sigma)}{\phi_y(\alpha,\alpha,M,\sigma)} < \frac{\phi_\alpha(\alpha,c_n,M,\sigma)}{\phi_y(\alpha,c_n,M,\sigma)}$ *for $y = M, \sigma$ and $c_n < \alpha$.*

(D) (convexity) $\int_{c_L}^{\alpha} \phi(\alpha, c_n, M, \sigma)\Gamma(c_n)\,dc_n < (\bar{c} - c_L)\phi(\alpha, c_L, M, \sigma) + (\alpha - \bar{c})\phi(\alpha, \alpha, M, \sigma)$, *where* $\Gamma(c_n) = \frac{g^\sigma(c_n) + (c_n - \bar{c})dg^\sigma(c_n)/dc_n}{g^\sigma(c_L)}$.

These assumptions are satisfied by the uniform distribution. Assumption 2.(A) states that the prior distribution is part of the family of bounded distributions $\mathcal{G}$, while Assumption 2.(B) is a monotonicity requirement. From Lemma (6), it also emerges that $\Phi(\alpha, M, \sigma)$, which represents the expected profit of a firm with cost equal to the cut-off point α, is decreasing in α and σ. Assumption 2.(C) is the single crossing condition: it guarantees the uniqueness of the equilibrium in the overall game. Assumption 2.(D) generally imposes a convexity-like condition, which is certainly verified if $\Gamma = 1$ (as in the uniform case) and ϕ is convex in c_n.

Under Assumption 2.(B), if $\Phi(c_H, M, \sigma) < 0$, then for every M, there exists an α^*, such that: $E_{G_{\alpha^*}}\left[\tilde{\Pi}_n|c_n, M\right] \lesseqgtr 0$ when $c_n \gtreqless \alpha^*$. Therefore, (16) the optimal entry threshold α^* is implicitly given by:

$$E_{G_{\alpha^*}}\left[\tilde{\Pi}_n|\alpha^*, M\right] = 0 \tag{16}$$

Lemma 7 *Given (16), under Assumption 2:*

$$\frac{d\alpha^*}{dM}<0;\ \frac{\partial\alpha^*}{\partial D}|_M>0;\ \frac{\partial\alpha^*}{\partial\sigma}|_M<0. \tag{17}$$

When $\Phi(c_H, M, \sigma) > 0$, $\alpha^* = c_H$ so that: $\frac{d\alpha^*}{dM} = 0$; $\frac{\partial\alpha^*}{\partial D}|_M = 0$ and $\frac{\partial\alpha^*}{\partial\sigma}|_M > 0$. Note how the second inequality of (17) seems to suggest that an increase in the market density D allows more inefficient firms to profitably stay in the market. However, this is not an equilibrium result for the full game because the partial derivatives in Lemma 7 are calculated for fixed M; hence they do not capture how a change in D affects the equilibrium number of firms in the pre-entry stage.

3.2 Licensing stage

The number of firms entering the licensing stage is obtained by assuming that firms correctly anticipate the outcomes of both the selection stage, i.e., based on the characteristics of the industry cost distribution, each firm expects that high-cost types may choose to exit the market, and the pricing stage.

Prior to paying the licensing fee F_L, firm n is unaware of its own type c_n, so it can base its entry decision only on the expected profit it would gain, which is a function of the endogenously determined number of entrants M into the licensing stage. All firms acquiring a licence enter the selection stage, but only if $c_n \leq \alpha^*$, firm n starts production after paying F_p, gaining $E_{G_{\alpha^*}}\left[\tilde{\Pi}_n|c_n, M\right]$; otherwise it stays out and gains 0.

From 15 and Assumption 2, the expected profit of a firm deciding to enter the licensing stage, gross of the entry fee F_L, depends, amongst other things, on the prior and posterior distributions of costs G and G_{α^*} and on the value of α^*:

$$E\left[\tilde{\Pi}_n|M\right] = \int_{c_L}^{\alpha^*} E_{G_{\alpha^*}}\left[\tilde{\Pi}_n|c_n, M\right] dG(c_n) = \int_{c_L}^{\alpha^*} \phi(\alpha^*, c_n, M, \sigma)\, dG(c_n) \tag{18}$$

Lemma 8 *Under Assumption 2, $E\left[\tilde{\Pi}_n|M\right]$ is monotonically decreasing in M.*

Therefore, we can use the zero-profit condition to determine the equilibrium number of entrants in the first stage, M_S:

$$E\left[\tilde{\Pi}_n|M_S\right] = F_L, \tag{19}$$

and $N_S = M_S G\left(\alpha^*\left(M_S\right)\right)$. Under Lemma 8 the solution is unique.

3.3 Main Results

The next two Propositions report comparative static results: the first constitutes the main thrust of this Section and shows new insights into the functioning of the selection mechanism by deriving predictions on the relationship between the truncation point of the prior distribution α^* and the variance of the same distribution, the latter being a measure of firm heterogeneity; the second analyzes the link between selection and market density.

Proposition 1 *Under Assumption 2:*

$$\frac{dM_S}{d\sigma} > 0; \ \frac{d\alpha^*}{d\sigma} \lesseqgtr 0 \ \textit{if} \ \Phi\left(c_H, M_S, \sigma\right) \lesseqgtr 0. \tag{20}$$

The first inequality confirms the result presented in Lemma 5.b), that cost variance increases expected profits and hence the number of potential entrants. Indeed, expected profits are convex in c_i so that, for the Jensen inequality, $E\left(\Pi(\tilde{c})\right) > \Pi(E(\tilde{c}))$ (Spulber, 1995). Thus, cost heterogeneity has a positive effect on M_S: as more potential professionals are attracted into the market, competition intensifies. This implies an increase in proximity among firms, and greater possibility for consumers to substitute one firm's services with another. Selection, and the associated exit of high-cost firms, takes place because these firms perceive they cannot compete adequately with low-cost ones, if the cost gap is sufficiently high and, therefore, they prefer to exit since, in the competition stage, it is unlikely that they will recover the second set-up sunk cost due to their low efficiency.

The selection mechanism induced by cost heterogeneity is formally illustrated in the second inequality. The cut-off point α^*, i.e., the level of cost at which the prior distribution is truncated, is not effective when the expected profit $\Phi\left(c_H, M_S, \sigma\right)$ of a firm with cost c_H is non-negative. That is, $\alpha^* = c_H$ if $\Phi\left(c_H, M_S, \sigma\right) > 0$.[11]

However, when selection is active, i.e. $\alpha^* < c_H$ because expected profits are negative at c_H, the cut-off point is negatively related to cost heterogeneity. Therefore, there exists

[11]Under Assumption 2, an increase in variance is obtained by keeping the ex-ante average cost constant and by moving mass from the center of the distribution to the tails. If G is symmetric, it can be obtained by increasing c_H and simultaneously reducing c_L of the same amount. If G is uniform $\sigma = \left(c_H - c_L\right)/2\sqrt{3}$, but when all the mass is equally concentrated on the extremes: $\sigma = \left(c_H - c_L\right)/2$.

a value of the cost gap $(c_H - c_L)$ beyond which a progressive truncation at $\alpha^* < c_H$ of the prior cost distribution occurs: the thrust of Proposition 1 is to show that the cut-off point α^* reduces as the cost gap goes beyond a certain level. The analysis thus bears important policy consequences: the selection mechanism becomes more severe precisely in those situations where it is more needed, that is, when there is a risk that highly inefficient firms may remain and operate in the market.

Proposition 2 *Under Assumption 2,*

$$\frac{d\alpha^*}{dD} < 0; \ \frac{dM_S}{dD} > 0; \ \frac{dM_S}{dF_L} < 0; \tag{21}$$

The first inequality mirrors the theoretical predictions in Syverson (2004): the upper bound of the firms' posterior cost distribution decreases in demand density. Part of the explanation of this result is associated with the second inequality: more firms enter the licensing stage in denser markets, attracted by a higher level of expected profit. Therefore, an increase in density plays a qualitatively similar role as cost variance. Because the average distance separating any two firms reduces, their products become closer substitutes and the ensuing intensification of competition makes it less likely for a high-cost firm to retain any positive market share. The posterior cost distribution obtains therefore from a truncation of the prior distribution from above, whose magnitude is larger in denser markets.

It is not possible, however, to derive a clear-cut prediction with regards to the impact of market density on the equilibrium number of firms, N_S. In this case, a direct and an indirect effect are at play:

$$\frac{dN_S}{dD} = \overbrace{\frac{dM_S}{dD}}^{>0} G(\alpha^*) + g(\alpha^*) \overbrace{\frac{d\alpha^*(M_S)}{dM}}^{<0} \overbrace{\frac{dM_S}{dD}}^{>0} M_S \gtreqless 0 \tag{22}$$

On the one hand, in a denser market larger profits intensify entry in the first stage, thereby also increasing the number of firms in the production stage (direct effect); on the other, the competitive pressure from a larger number of firms push towards the exclusion of less competitive firms (indirect effect). The overall effect depends on the net balance between these two forces. Simulations obtained assuming a uniform distribution suggest that the first effect dominates: N_S is increasing in D but at a lower rate than M_S.

Finally, the third inequality constitutes a standard result, which has however important implications in liberal professions' markets. On the one hand, a higher licensing fee discourage potential entrants, thereby creating a potential restriction to entry and therefore to competition; on the other, the licensing sunk cost plays a crucial role and its elimination would be likely associated with efficiency losses. For instance, a strictly positive value has been shown to alleviate moral hazard problems associated with the provision of high-quality services (Shapiro, 1986; Kleiner, 2000). In addition to this positive outcome, Proposition 1 shows that the licensing stage plays another important role in the literature on professional markets: it induces self-selection of most capable professionals in situations where heterogenity is large. Therefore, licensing plays a double role guaranteeing both the quality of professional services and their long-run efficiency.

Figure 1 further illustrates how the selection mechanism operates as a function of the cost gap (which is a proxy for variance) $c_H - c_L$.[12]

For low levels of the cost gap, the optimal number of firms in the licensing stage and in the production stage coincide ($M_S = N_S$ and $\alpha^* = c_H$). That is, there is no selection when heterogeneity is small in magnitude. However, as heterogeneity increases, the selection mechanism determines a cut-off point after which $\alpha^* < c_H$ and M_S increases at a fast rate. When the decision to enter is made under uncertainty, an increase in the costs' variance leads to an increase in the expected profit in the licensing stage, even when firms can anticipate that in the subsequent selection stage a number of firms will not continue. In Figure 2, the selection mechanism is responsible for: 1) the decreasing trend in N_S due to α^* being decreasing in cost variance when selection is in place; 2) the exit of the $M_S - N_S$ least efficient firms ($M_S - N_S$ also increases with variance); 3) the increase in the average level of market efficiency, $\bar{c}_e = \frac{\alpha^* - c_L}{2} < \bar{c}$.

To sum up, the foregoing analysis, in addition to being able to replicate some recent results explaining the relatively higher efficiency of firms operating in denser markets, has illustrated how the licensing stage may have beneficial implications for allocative efficiency even in the absence of the traditional rationales generally advocated to justify licensing as a regulatory instrument in liberal professions (Philipsen, 2009). Indeed, our set-up abstracts from any form of market failure due to informational asymmetries, negative externalities or the provision of a public good (Kleiner, 2000). Nonetheless, we obtain that the licensing stage is instrumental in allowing a selection mechanism to operate by driving inefficient firms out of the market. Most importantly, our main prediction is that such a mechanism is more restrictive in markets characterized by higher levels of ex-ante cost heterogeneity. This may be a possible reasons why licensing is required for the exercise of some liberal professions (lawyers) but not for others (pharmacists). The next Section focuses on the latter markets and investigates how cost heterogeneity affects their long-run

[12] The numerical values in Figure 1 were obtained assuming an uniform distribution, and parameters' values: $F_L = 0.2$, $F_p = 0.8$, $\bar{c} = 0.1$, $c_H - c_L \in [0, 0.14]$, $D = 100$ and $t = 1$.

market structure configuration.

4 Entry without licensing

For some professions, acquiring a relevant university degree is the only formal condition to start production. For instance, as Schaumans and Verboven (2008) illustrate, in Belgium all individuals gaining a degree in pharmaceutical sciences do not need to satisfy any further educational requirement and are consequently deemed qualified to run an establishment in any part of the country. There is, therefore, no licensing stage involved. However, in Belgium and by different degrees in many other EU countries, the number of pharmacies allowed to operate within a given geographic area is fixed and depends on the area population.

Quite interestingly, when licensing is required, then entry restrictions are not applied. For instance, in Belgium medical graduates have to study two further years to become 'general practitioners', and five years to become 'specialists' (Schaumans, 2009). There are however no regulatory entry restrictions on the number of enterprises run by doctors in any country. It appears therefore that licensing and entry restrictions constitute mutually exclusive regulatory mechanisms, the presence of the latter being generally explained by the public interest motive according to which unregulated markets would generate excessive entry and therefore a lower level of social welfare (Philipsen, 2009). In this Section we build on the foregoing analysis to consider the implications of applying entry restrictions in markets where licensing is not a legal requirement.

We consider a two-stage entry game with the entry stage followed by the production stage. This is directly obtained from the licensing entry game of the previous Section by setting $F_L = 0$, i.e., no licensing. Before entering, each potential entrant only knows the costs distribution but has no information on the actual realization of its own and its potential opponents' costs, as in Assumption 1. In the first stage, those potential entrants choosing to enter have to pay a set-up cost F, the others remain out of the market and gain zero profit. After investing F and occupying a random equidistant location, in the production stage price competition takes place among the N entrants as in Section 2, so that locations and costs are common knowledge. We solve the model by backward induction. The price equilibrium in the production stage is therefore given by Lemma 2 under the existence conditions of Lemma 3.

4.1 Free market entry equilibrium

Firms are risk-neutral and have perfect foresight, so that they enter the market if their expected profit is non-negative. Since costs are unknown in the first stage, firm n's random profit is:

$$\tilde{\Pi}_n = \frac{D}{t}\left(\tilde{p}_n - \tilde{c}_n\right) \cdot \left(\frac{1}{2}\left(\tilde{p}_{\langle n-1\rangle} + \tilde{p}_{\langle n+1\rangle}\right) - \tilde{p}_n + \hat{k}\right) - F. \tag{23}$$

Lemma 9 *The expected profit of firm n is:*

$$E\left[\tilde{\Pi}_n \left| N\right.\right] = \frac{D}{t}\left(\sigma^2 W_I(N) + \hat{k}^2\right) - F \tag{24}$$

- $W_I(N) = \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|}\left[\left(w_{\bar{l}} - w_{\bar{l}-1}\right) - \left(w_{\bar{l}-|i|} - w_{\bar{l}-|i+1|}\right)\right] > 0$, *with* $w_{-1} := w_{\bar{l}+\underline{l}}$*;*
- $\lim_{N\to\infty} W_I(N) = \bar{W}_I = 1 - \frac{4}{9}\sqrt{3} \simeq 0.23020.$

The values of $W_I(N)$ are reported in Table 1: they monotonically decrease for $N \geq 3$, and rapidly converge toward $\bar{W}_I$. From Lemma 9, the expected profit $E\left[\tilde{\Pi}_n \left| N\right.\right]$ decreases in N and F and increases in D, σ and t.[13] As in the licensing game, the expected profits increase with cost heterogeneity. Because the profit function is convex in c_i, in the extreme case where any two adjacent firms' costs are perfectly negatively correlated, i.e., $E\left[\left(c_n - \bar{c}\right) c_{\langle n+1\rangle}\right] = -\sigma^2$ and N is even, it can be shown that the expected profit (24) reduces to $E[\Pi] = \frac{D}{t}\left(\hat{k}^2 + \frac{4}{9}\sigma^2\right) - F$.[14] That is, firms prefer to "gamble" even if there is a fifty-fifty chance of drawing a high cost. Furthermore, when there is perfect positive pairwise correlation among costs, $\sigma^2 = 0$ and the profits are the same as the case of no uncertainty and no heterogeneity.

Proposition 3 *From Lemma 9, the equilibrium number of firms in the market,* N_M*, is implicitly given by the following equation:*

$$D \cdot \left(\frac{t}{N_M^2} + \frac{\sigma^2}{t} W_I(N_M)\right) = F$$

[13] Taking the derivative of $E\left[\tilde{\Pi}_n \left| N\right.\right]$ with respect to t, we obtain that the expected profit is increasing in t when $\sigma < \hat{k}/\sqrt{W_I} \simeq 2.09\hat{k}$. In the support $[c_L, c_H]$ the maximal variance arises from a distribution where all the mass is equally concentrated on the extremes: $\sigma = (c_H - c_L)/2$; hence, under (9) the expected profit is always increasing in t for any distribution.

[14] Perfect negative correlation is obtained when the 1-step neighbors of a c_L-type firm are c_H-types, and viceversa. Calculations are available on request.

For N large, the number of firms is approximated by:

$$N_M \simeq \sqrt{\frac{t}{F/D - \bar{W}_I \sigma^2/t}} \tag{25}$$

It is noteworthy that N_M converges to the value in Salop (1979) when firms' costs are identical, i.e., $N_M = \sqrt{tD/F}$ when $\sigma^2 = 0$. More importantly, the game with cost heterogeneity yields a higher number of firms in equilibrium, relative to that in the traditional case. However, the two cases are not perfectly comparable given the different informational structure they assume. Therefore, to gauge the extent by which the free-entry outcome in the last Proposition is sub-optimal, we need to derive the Pareto optimal outcome under cost heterogeneity.

4.2 Pareto Efficient Entry

In the spirit of Salop (1979), in this Section we compare the free-entry market equilibrium in Proposition 3 with the first-best optimum that would be chosen by a social planner. It is worth stressing that relative to the welfare analysis in Salop (1979), introducing cost heterogeneity brings about an extra distortion associated with the total production cost incurred in the economy. This is captured by the difference between the market equilibrium prices in Lemma 1 and the first-best prices in Lemma 4. Indeed, the market equilibrium pricing rule induces an allocative inefficiency such that consumers who should patronize a low-cost firm end up buying from a higher cost firm, thereby increasing the total and average costs incurred in the industry. It follows therefore that the first-best optimal number of firms chosen by a benevolent social planner is such that it minimizes the expected sum of transport, production and fixed costs (see Appendix, proof of Lemma 4).

Proposition 4 *a) The first-best optimal number of firms is:*

$$N_F = \frac{1}{2}\sqrt{\frac{t}{F/D - 0.5\sigma^2/t}}. \tag{26}$$

b) There is less excessive entry under cost heterogeneity, because $\bar{W}_I < 0.5$, so that $N_M/N_F < 2$ when $\sigma > 0$.

Part b) of the Proposition refers to the fact that the ratio $N_M/N_F = 2$ when firms are identical. Its corollary is that the greater the degree of heterogeneity, the less excessive market entry is, with the proviso that condition (9) determines a lower bound for N_M/N_F.

The welfare-enhancing effect of cost heterogeneity could be seen as unexpected, given the extra allocative distortion that the social planner has to manage in this case. From Proposition 3 and eq. (41) in the Appendix, it appears clear that an increase in variance raises a firm's expected profit less that it reduces the industry's expected total cost (in particular, the average variable cost). Therefore, the regulator is willing to allow more firms in the market, so as to shift production towards the more efficient plants, since the gain from doing so offsets the increase in expected transport cost. In other words, excessive entry is reduced because the first-best number of firms, N_F, increases with cost heterogeneity faster than the free-entry one, N_M.

The overall implications of the foregoing analysis on the regulation of liberal professions point in opposite directions. On the one hand, when we consider the more realistic case of heterogeneous firms, the free market outcome tends to be more closely aligned to the first-best equilibrium, thereby suggesting there is less need to restrict the number of firms in professional markets with more heterogeneity. On the other, in the model the distance between the two outcomes remains sufficiently large. Indeed, numerical simulations show that N_M/N_F is indeed monotonically decreasing in σ and reaches a minimum value around 1.78, when $(c_H - c_L)$ in Lemma 3 takes its maximum possible value.[15] Therefore, the motivation for entry restrictions on the maximum number of firms advocated in the public interest view appears to continue to be theoretically relevant also in our setting.

4.3 Discussion

To further highlight the welfare-enhancing role of cost heterogeneity, consider the following three different pricing schemes:

- *uniform* (e.g., fixed or minimum prices set by professional bodies correspond to this case; equivalently, the regulator may not differentiate between different types of suppliers, as in the case of the Belgian pharmacies described in Schaumans and Verboven (2008), where urban and rural pharmacies are compensated uniformly);
- *market* (prices are set according to Lemma 2);
- *first-best* (the marginal cost pricing in (10)).

Each of these is characterized by a different expected average social cost, $E[C(N)]$, given by the sum of the expected average transport, production and fixed costs. Analytically, and without loss of generality, we have

$$E[C(N)] = \frac{4\bar{c} + \hat{k}}{4} + \frac{NF}{D} - \frac{\sigma^2}{2\hat{k}}\chi, \tag{27}$$

[15] The other values were: $F = t = 1$, $D = 100$, $\bar{c} = 0.1$.

where $\chi = 0$ under uniform pricing, $\chi \simeq 0.65$ under market pricing, and $\chi = 1$ under first-best pricing - see (41) in the Appendix for the first-best case. Note that the cost associated to uniform pricing is not affected by σ, and that three costs coincide when $\sigma = 0$.[16]

Figure 2 plots $E[C(N)]$ under the three pricing scenarios as a function of the number of firms. It depicts the relative position of a number of possible outcomes: F is the first-best solution; S is the second best solution (the social planner's choice of entrants when they set market prices); M is the free-market outcome; and U is the social planner's choice of entrants under uniform pricing. Assuming that the first-best solution F cannot be implemented (e.g., due to lack of information about cost heterogeneity), policy intervention should be determined on the basis of the optimal mix between the possible price setting alternatives and the related entry restrictions.

Figure 2 unambiguously shows that uniform price regulation is dominated by the free-market pricing outcome, and that it performs very poorly when heterogeneity is taken into account.[17] With regards to entry, the free market equilibrium in M, defined in Proposition 3, tends to be less inefficient than U, suggesting that some form of entry regulation may be needed. Indeed, the combination of free market prices with entry restrictions on the number of firms (point S) constitutes a form of second-best, since it entails a level of $E[C(N)]$ only slightly above that of F, but is relatively easier and less costly to implement than the first-best outcome since it does not need information on firms' costs.

To conclude, in the presence of sufficiently large cost heterogeneity, welfare improvements can be achieved by replacing uniform pricing with a policy combining both entry restrictions and a price mechanism that more closely reflects firms' underlying costs. Our analysis therefore lends theoretical support and is largely consistent with some policy recommendations found in the empirical literature. In their counterfactual analysis on the Belgian pharmacies, Schaumans and Verboven (2008) consider two policy measures and find that if the Belgian authorities: 1) relaxed the restriction on entry by allowing twice as many outlets per given area population size, and at the same time 2) they cut the regulated prices non-uniformly by favoring the pharmacies in smaller markets (which are likely to have higher costs), then the number of pharmacies and the related geographical coverage would remain very similar, if not slightly higher, than the pre-intervention period. Therefore, the combined measures implicitly describe a regulatory intervention that

[16] They differ due to the impact that each pricing mechanism has on market shares, which in turn affects the expected transport and production costs, as the proof of Proposition 4 in the Appendix highlights.

[17] In a study of the U.S. real estate market, where commissions rates are fixed at the 6% level and barriers to entry are very low, Hsieh and Moretti (2003) report evidence indicating that fixed rates lead to highly wasteful entry.

would move the Belgian pharmacies market from point U (i.e., fewer firms and uniform prices) towards points F and S (slight increase in firms' number and prices reflecting underlying cost conditions) in Figure 2.[18] Based on our analysis, such a move would be welfare-enhancing.

5 Conclusions

Professional services are a key sector in many modern economies. Such a sector is characterized, however, by high levels of regulation, resulting from a mix of State regulation, self-regulation and custom and practice. In European Commission (2004), all EU member States and competition authority were invited to assess to what extent existing professional regulations and rules truly serve the public interest and can be objectively justified. To this purpose, the Commission suggested the application of a "proportionality test" according to which each regulatory measure should be explained by an explicitly stated objective detailing how the chosen measure represents the least restrictive mechanism of competition to effectively attain the stated objective. The Commission's recommendation implicitly calls for further analysis on the economic properties associated with each regulatory practice.

In this paper we have investigated the economic implications of regulations pertaining both to entry requirements in professional markets and to price setting. Licensing appears to be particularly important in a market with heterogenous firms, since it is capable to induce a mechanism such that only the most efficient firms in the market become operative. Such a selection mechanism intensifies its effect as the degree of heterogeneity increases. Our findings therefore point at a generally beneficial effect of licensing, and may explain why a formal licensing law capturing many features present in many Western World economies, has been recently introduced in China for the legal professions (Philipsen, 2009). Most importantly, this beneficial role of licensing is not associated with the need to remedy the negative effects of asymmetric information between customer and service supplier.

For relatively simpler professional tasks, it is often argued that licensing may constitute an excessively restrictive rule, which, if relaxed, might benefit consumers. For example, prices dropped significantly in the United Kingdom after the lawyers' reserved rights to provide conveyancing services were removed (European Commission, 2004). We investigate the extreme case of no licensing, to find that when the first-best outcome cannot be implemented, the second-best policies should be aimed at combining the greatest possible

[18] Because market density is fixed at D in both cases, the number of firms per capita is therefore larger under SB.

flexibility in price setting, so that prices reflect the underlying efficiency of each firm, with some limits on the number of practitioners in the market. From a practical viewpoint, our analysis suggests, in line with the evidence available in the literature, that a reform removing uniform pricing in favor of a liberalized price setting system may also likely necessitate a re-adjustment of the restrictive rules on entry based on a geographic market's demographics.

Appendix

Proof for Lemma 1. See Alderighi and Piga (2008).

Proof for Lemma 2. While a more general proof can be obtained on request, here we provide the intuition behind the derivation of (2) and (8) assuming N is even. From Lemma 1, firm 0's equilibrium strategy is of the form:

$$p_0 = w_{\bar{l}} c_0 + w_{\bar{l}-1} (c_{-1} + c_{+1}) + w_{\bar{l}-2} (c_{-2} + c_{+2}) + ... + w_0 c_{\bar{l}} + k. \tag{28}$$

By the same token, using the shift in weights' notation described in Figure 1, the price charged by the adjacent firms located at -1 and $+1$, respectively, are:

$$\begin{aligned} p_{-1} &= w_{\bar{l}} c_{-1} + w_{\bar{l}-1} (c_{-2} + c_0) + w_{\bar{l}-2} (c_{-3} + c_{+1}) + ... + w_0 c_{\bar{l}-1} + k & (29) \\ p_{+1} &= w_{\bar{l}} c_{+1} + w_{\bar{l}-1} (c_0 + c_{+2}) + w_{\bar{l}-2} (c_{-1} + c_{+3}) + ... + w_0 c_{\underline{l}} + k. & (30) \end{aligned}$$

Now substituting (28), (29) and (30) in (2), with $n = 0$, and collecting the similar terms, we obtain:

$$\begin{aligned} & c_0 \left(4w_{\bar{l}} - 2 - 2w_{\bar{l}-1}\right) + (c_{-1} + c_1) \left(4w_{\bar{l}-1} - w_{\bar{l}-2} - w_{\bar{l}}\right) + (c_{-2} + c_2) \left(4w_{\bar{l}-1} - w_{\bar{l}} - w_{\bar{l}-2}\right) + .. \\ & .. + \left(c_{\underline{l}} + c_{\bar{l}-1}\right) (4w_1 - w_0 - w_2) + c_{\bar{l}} (4w_0 - 2w_1) + \left(2k - 2\hat{k}\right) = 0. \end{aligned} \tag{31}$$

In order for this equality to hold for every possible configuration of c_n, $n \in \mathbb{L}$, all the expressions in round brackets must be equal to zero, thereby proving (4)-(6) and (8). Finally, (7) emerges by replacing $c_n = c$ for $n \in \mathbb{L}$ in (31).■

Proof for Lemma 3. See Alderighi and Piga (2008).

Proof of Lemma 4.

Assume there are N firms already settled on the circle, with a generic firm n having cost $c_n \in [c_L, c_H]$. We focus on the generic arc of length $\frac{1}{N}$ defined by the locations of firms $\langle n-1 \rangle$ and n. Because demand is inelastic, the first-best allocative solution is reached when the sum of production and transport costs is minimized. The solution to this problem, which is worked out when the planner has learnt the cost realization in each location, is equivalent to finding the optimal allocation of consumers between two adjacent firms. However, unlike the standard case, where the planner splits the arc evenly among firms, optimality under cost heterogeneity may require that the most efficient firms supply a larger proportion of the market even if this increases transport costs.

Let d_n and $d_{\langle n-1\rangle} = \left(\frac{1}{N} - d_n\right)$ be the distance of the pivotal consumer from two adjacent firms. The average distance between a generic consumer on the arc and the seller she patronizes is:

$$\bar{d} = N\left(\int_0^{d_{\langle n-1\rangle}} xdx + \int_0^{d_n} xdx\right) = \frac{N}{2}\left(d_{\langle n-1\rangle}^2 + d_n^2\right). \tag{32}$$

Hence, the average transport costs among all consumers on the arc is:

$$TC = \bar{d}t = \frac{N}{2}\left(d_{\langle n-1\rangle}^2 + d_n^2\right)t \tag{33}$$

The proportion of consumers patronizing firm $\langle n-1\rangle$ and firm n are, respectively, $Nd_{\langle n-1\rangle}$ and Nd_n. The average (variable) cost of producing goods for a consumer on the arc is:

$$VC = Nd_{\langle n-1\rangle}c_{\langle n-1\rangle} + Nd_nc_n \tag{34}$$

The social planner problem can be expressed as:

$$\min_{d_n\geq 0}(TC+VC) = \frac{N}{2}\left(\left(\frac{1}{N} - d_n\right)^2 + d_n^2\right)t + N\left(\frac{1}{N} - d_n\right)c_{\langle n-1\rangle} + Nd_nc_n \tag{35}$$

Therefore,

$$d_n^* = \frac{1}{2N} + \left(\frac{c_{\langle n-1\rangle} - c_n}{2t}\right). \tag{36}$$

Its implementation can be obtained by using the prices in the Lemma as policy instruments. ■

Proof of Lemma 5. Multiply the terms within the square brackets in (11) to obtain: $E\left[\tilde{p}_n\tilde{p}_{\langle n-1\rangle}\right] - E\left[\tilde{p}_n^2\right] + E\left[\tilde{p}_n\right]\hat{k} - c_nE\left[\tilde{p}_{\langle n-1\rangle}\right] + c_nE\left[\tilde{p}_n\right] - c_n\hat{k}$. Each term is derived separately.

Part 5.I *Under Assumption 1, if firm n knows its costs but not those of its opponents and in the final stage prices follow Lemma 2, then*
$E_{G_e}\left[\tilde{p}_{\langle n\rangle}^2\right] = \left(\left(w_{\bar{l}}c_n + \left(1 - w_{\bar{l}}\right)\bar{c}_e\right) + \hat{k}\right)^2 + \sigma_e^2W_{L.0}(N)$, *where* $W_{L.0}(N) = \sum_{i=\mathbb{L}_0}\left(w_{\bar{l}-|i|}\right)^2$.

Derivation of Part 5.I. To save notation, we do not explicitly show the dependence of the expected value on G_e. Using (3), $E\left[c_n\right] = c_n$, $E\left[\tilde{c}_{\langle n+i\rangle}\right] = \bar{c}_e, E\left[c_n^2\right] = c_n^2$, $E\left[c_n\tilde{c}_{\langle n+i\rangle}\right] = c_n\bar{c}_e$, $E\left[\tilde{c}_{\langle n+i\rangle}^2\right] =$

$(\sigma^2+\bar{c}_e^2)$, with $i\in\mathbb{L}_0$ and $E\left[\tilde{c}_{\langle n+i\rangle}\tilde{c}_{\langle n+j\rangle}\right]=\bar{c}^2$, $i,j\in\mathbb{L}_0$, $i\neq j$:

$$\begin{aligned} E\left[\tilde{p}_{\langle n\rangle}^2\right]=&E\left[\left(w_{\bar{l}}\cdot c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\cdot\tilde{c}_{\langle n+i\rangle}+\hat{k}\right)^2\right]\\ &E\left[\left(w_{\bar{l}}\cdot c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\cdot\tilde{c}_{\langle n+i\rangle}\right)^2\right]+2\left(w_{\bar{l}}\cdot c_n+\left(1-w_{\bar{l}}\right)\bar{c}_e\right)\hat{k}+\hat{k}^2. \end{aligned} \tag{37}$$

The expected value in the squared brackets can be written as: $w_{\bar{l}}^2c_n^2+2w_{\bar{l}}\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\bar{c}_ec_n+\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}^2E\left[\tilde{c}_{\langle n+i\rangle}^2\right]+\sum_{i\neq j;i,j=\mathbb{L}_0}w_{\bar{l}-|j|}w_{\bar{l}-|i|}E\left[\tilde{c}_{\langle n+i\rangle}\tilde{c}_{\langle n+j\rangle}\right]$, or:

$$w_{\bar{l}}^2c_n^2+2w_{\bar{l}}\left(1-w_{\bar{l}}\right)\bar{c}_ec_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}^2\sigma_e^2+\sum\nolimits_{i,j=\mathbb{L}_0}w_{\bar{l}-|j|}w_{\bar{l}-|i|}\bar{c}_e^2.$$

Setting $W_{L.0}(N)=\sum_{i=\mathbb{L}_0}\left(w_{\bar{l}-|i|}\right)^2$ and noting that $\sum_{i,j=\mathbb{L}_0}w_{\bar{l}-|j|}w_{\bar{l}-|i|}=\left(1-w_{\bar{l}}\right)^2$, we obtain $\left(w_{\bar{l}}c_n+\left(1-w_{\bar{l}}\right)\bar{c}_e\right)^2+\sigma_e^2W_{L.0}(N)$. Summing up previous results completes the calculation. $\diamondsuit$

Part 5.II *Under Assumption 1, if firm n knows its costs but not those of its opponent and in the final stage prices follow Lemma 2 then*
$E_{G_e}\left[\tilde{p}_{\langle n\rangle}\tilde{p}_{\langle n-1\rangle}\right]=\left(c_nw_{\bar{l}-1}+\left(1-w_{\bar{l}-1}\right)c_e+\hat{k}\right)\left(c_nw_{\bar{l}}+\left(1-w_{\bar{l}}\right)c_e+\hat{k}\right)+W_{L.1}(N)\sigma^2$, *where* $W_{L.1}(N)=\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}w_{\bar{l}-|i+1|}$.

Derivation of Part 5.II. The proof is similar to that of Part 5.I. Using (**3**), we obtain: $E\left[\tilde{p}_{\langle n\rangle}\tilde{p}_{\langle n-1\rangle}\right]=E\left[\left(w_{\bar{l}}\cdot c_n+\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\cdot\tilde{c}_{\langle n+i\rangle}+\hat{k}\right)\left(w_{\bar{l}-1}\cdot c_n+\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i+1|}\cdot\tilde{c}_{\langle n+i\rangle}+\hat{k}\right)\right]$. This can be written as:

$$\begin{aligned} &E\left[\left(w_{\bar{l}}c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\tilde{c}_{\langle n+i\rangle}\right)\left(w_{\bar{l}-1}c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i+1|}\tilde{c}_{\langle n+i\rangle}\right)\right]+\\ &+\left(w_{\bar{l}}c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}\bar{c}_e\right)\hat{k}+\left(w_{\bar{l}-1}c_n+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i+1|}\tilde{c}_{\langle n+i\rangle}\right)\hat{k}+\hat{k}^2.\text{ Or:} \end{aligned}$$

$$\begin{aligned} &w_{\bar{l}}w_{\bar{l}-1}c_n^2+w_{\bar{l}}\left(1-w_{\bar{l}-1}\right)c_n\bar{c}_e+w_{\bar{l}-1}\left(1-w_{\bar{l}}\right)c_n\bar{c}_e+\sum\nolimits_{i=\mathbb{L}_0}w_{\bar{l}-|i|}w_{\bar{l}-|i+1|}\left(\sigma^2+\bar{c}_e^2\right)\\ &+\sum\nolimits_{i\neq j;i,j=\mathbb{L}_0}w_{\bar{l}-|i|}w_{\bar{l}-|j+1|}\bar{c}_e^2+\left(w_{\bar{l}}\cdot c_n+\left(1-w_{\bar{l}}\right)\bar{c}_e\right)k+\left(w_{\bar{l}-1}c_n+\left(1-w_{\bar{l}-1}\right)\bar{c}_e\right)\hat{k}+\hat{k}^2. \end{aligned}$$

Noting that $\sum_{i,j=\mathbb{L}_0}w_{\bar{l}-|i|}w_{\bar{l}-|j+1|}=\left(1-w_{\bar{l}}\right)\left(1-w_{\bar{l}-1}\right)$, after some simplifications:

$$\begin{aligned} &w_{\bar{l}}w_{\bar{l}-1}c_n^2+w_{\bar{l}}\left(1-w_{\bar{l}-1}\right)c_n\bar{c}_e+w_{\bar{l}-1}\left(1-w_{\bar{l}}\right)c_n\bar{c}_e+\left(1-w_{\bar{l}}\right)\left(1-w_{\bar{l}-1}\right)\bar{c}_e^2+W_{L.1}(N)\sigma^2+\\ &+\left(w_{\bar{l}}\cdot c_n+\left(1-w_{\bar{l}}\right)\bar{c}_e\right)\hat{k}+\left(w_{\bar{l}-1}c_n+\left(1-w_{\bar{l}-1}\right)\bar{c}_e\right)\hat{k}+\hat{k}^2, \end{aligned}$$

with: $W_{L.1}(N)=\sum_{i=\mathbb{L}_0}w_{\bar{l}-|i|}w_{\bar{l}-|i+1|}$. The thesis directly follows from factorization. $\diamondsuit$

a) To derive (13), substitute $E\left[\tilde{p}_n\right] = w_{\bar{l}}c_n + \left(1 - w_{\bar{l}}\right)\bar{c}_e + \hat{k}$, $E\left[\tilde{p}_{\langle n-1\rangle}\right] = w_{\bar{l}-1}c_n + \left(1 - w_{\bar{l}-1}\right)\bar{c}_e$ $+\hat{k}$ and the results in Parts 5.I and 5.II into (11):

$$
\begin{aligned}
& E_{G_e}\left[\tilde{\Pi}_n | c_n, N, M\right] = \\
& \left(c_n w_{\bar{l}-1} + \left(1 - w_{\bar{l}-1}\right)\bar{c}_e + \hat{k}\right)\left(c_n w_{\bar{l}} + \left(1 - w_{\bar{l}}\right)\bar{c}_e + \hat{k}\right) + W_{L.1}\sigma^2 \\
& -\left(\left(w_{\bar{l}}c_n + \left(1 - w_{\bar{l}}\right)\bar{c}_e\right) + \hat{k}\right)^2 - \sigma_e^2 W_{L.0} + \left(w_{\bar{l}}c_n + \left(1 - w_{\bar{l}}\right)\bar{c}_e + \hat{k} - c_n\right)\hat{k} \\
& -c_n\left(w_{\bar{l}-1}c_n + \left(1 - w_{\bar{l}-1}\right)\bar{c}_e + \hat{k}\right) + c_n\left(w_{\bar{l}}c_n + \left(1 - w_{\bar{l}}\right)\bar{c}_e + \hat{k}\right) \\
= \; & \left(\hat{k} + \bar{c}_e - c_n - c_e w_{\bar{l}} + c_n w_{\bar{l}}\right)\left(\hat{k} - \bar{c}_e w_{\bar{l}-1} + c_n w_{\bar{l}-1} + \bar{c}_e w_{\bar{l}} - c_n w_{\bar{l}}\right) + \left(W_{L.1} - W_{L.0}\right)\sigma_e^2 \\
= \; & \left(\left(1 - w_{\bar{l}}\right)\left(\bar{c}_e - c_n\right) + \hat{k}\right)\left(\left(w_{\bar{l}} - w_{\bar{l}-1}\right)\left(\bar{c}_e - c_n\right) + \hat{k}\right) + W_L(N)\,\sigma_e^2 \\
= \; & \left(w_d\left(\bar{c}_e - c_n\right) + \hat{k}\right)^2 + W_L(N)\,\sigma_e^2
\end{aligned}
$$

where $w_d = \left(1 - w_{\bar{l}}\right) = \left(w_{\bar{l}} - w_{\bar{l}-1}\right)$ and
$W_L(N) = \left(W_{L.1}(N) - W_{L.0}(N)\right) = \sum_{i=\mathbb{L}_0} w_{\bar{l}-|i|}\left(w_{\bar{l}-|i+1|} - w_{\bar{l}-|i|}\right) > 0.$

To compute $\bar{W}_L$, note that for N large enough we can use the numerical algorithm for the weights in fn. 7. Hence $W_L(N) \simeq \sum_{i=\underline{l}}^{\bar{l}}\left(yz^{|i|} \cdot yz^{|i+1|}\right) - \sum_{i=\underline{l}}^{\bar{i}}\left(yx^{-|i|}\right)^2 - y\left(yx^{-1} - y\right) = y^2\frac{x^{-1}-1}{x^{-1}+1} - y^2\left(x^{-1} - 1\right) = \frac{y^2}{x}\frac{1-x^{-1}}{1+x^{-1}} = \frac{2}{9}\sqrt{3} - \frac{1}{3} = 0.05156\,7.$

b) For (13) to be decreasing in c_n, it sufficient that $\left(w_d\left(\bar{c}_e - c_n\right) + \hat{k}\right) \geq 0$. First, for every G_e, $\left|\bar{c}_e - c_n\right| \leq \left|c_H - c_L\right|$. Hence $\left(w_d\left(\bar{c}_e - c_n\right) + \hat{k}\right) \geq 0$ holds if $\left|c_H - c_L\right| < \frac{1}{w_d}\hat{k}$, which is satisfied by (9). ■

Proof of Lemma 6. Omitted. ■

Proof of Lemma 7. First inequality. By totally differentiating (16) with respect to α^* and M, we obtain: $\frac{d\alpha^*}{dM} = -\Phi_M/\Phi_\alpha$ which is negative since $\Phi_M < 0$ from Lemma 6, and $\Phi_\alpha < 0$ from Assumption 2. The second and the third inequalities are derived using similar arguments. ■

Proof of Lemma 8. Note that $\frac{d}{dM}E_{G_{\alpha^*(M)}}\left[\tilde{\Pi}_n|, M\right] = \frac{d\alpha^*(M)}{dM} \cdot \Phi\left[\alpha^*(M), M\right]G_\alpha\left(\alpha^*(M)\right) + \int_{c_L}^{\alpha^*(M)}\frac{d}{dM}\left(\phi\left(\alpha^*(M), c_n, M, \sigma\right)\right)dG\left(c_n\right)$. Since the first addendum is null due to the zero-profit condition (16), a necessary and sufficient condition for monotonicity is that $\int_{c_L}^{\alpha}\frac{d}{dM}\phi\left(\alpha, c_n, M, \sigma\right) < 0$. However a sufficient condition is:

$$
\frac{d}{dM}\phi\left(\alpha^*(M), c_n, M, \sigma\right) = \phi_\alpha\left(\alpha^*(M), c_n, M, \sigma\right)\frac{d\alpha^*}{dM} + \phi_M\left(\alpha^*(M), c_n, M, \sigma\right) < 0, \tag{38}
$$

for any $c_n \leq \alpha$. Because $\phi_M < 0$ from Lemma 6 and $\phi_\alpha < 0$ from Assumption 2.(B), then $\phi\left(c_H, c_H, M, \sigma\right) > 0$ implies $\frac{d\alpha^*}{dM} = 0$, and (38) is satisfied. When $\phi\left(c_H, c_H, M, \sigma\right) < 0$, $\frac{d\alpha^*}{dM} = -\frac{\phi_M}{\phi_\alpha + \phi_{c_n}}\left(\alpha^*(M), \alpha^*(M), M, \sigma\right)$. Plugging the previous equation in (38) and dividing by

$-\phi_\alpha(\alpha^*(M), c_n, M, \sigma)$, we obtain:

$$\frac{\phi_M(\alpha^*(M), \alpha^*(M), M, \sigma)}{\phi_\alpha(\alpha^*(M), \alpha^*(M), M, \sigma) + \phi_{c_n}(\alpha^*(M), \alpha^*(M), M, \sigma)} < \frac{\phi_M(\alpha^*(M), c_n, M, \sigma)}{\phi_\alpha(\alpha^*(M), c_n, M, \sigma)}, \tag{39}$$

which holds thanks to Assumption 2.(C). ■

Proof of Proposition 1. First inequality - Totally differentiating (19), and recalling that from Lemma 8: $\frac{d}{dM} E_{G^\sigma_{\alpha^*(M)}}\left[\tilde{\Pi}_n|, M\right] < 0$, it follows that:
$sign\left[\frac{dM_S}{d\sigma}\right] = sign\left[\frac{d}{d\sigma} E_{G^\sigma_{\alpha^*(M)}}\left[\tilde{\Pi}_n|, M\right]\right]$. Note that:

$$\frac{d}{d\sigma}\int_{c_L}^{\alpha^*}\left(\phi\left(\alpha^*, c_n, M, \sigma\right)\right) g(c_n)\, dc_n =$$
$$\frac{d\alpha^*}{d\sigma}\cdot\phi\left(\alpha^*, \alpha^*, M, \sigma\right) g^\sigma\left(\alpha^*\right) - \frac{dc_L}{d\sigma}\cdot\phi\left(\alpha^*, c_L, M, \sigma\right) g^\sigma(c_L) + \int_{c_L}^{\alpha^*}\frac{d}{d\sigma}\left(\phi\left(\alpha^*, c_n, M, \sigma\right) g^\sigma(c_n)\right) dc_n,$$

where $\phi(\alpha^*, \alpha^*, M, \sigma) = 0$. After expanding the last term:

$$-\frac{dc_L}{d\sigma}\phi\left(\alpha^*, c_L, M, \sigma\right) g^\sigma(c_L) + \int_{c_L}^{\alpha^*}\phi\left(\alpha^*, c_n, M, \sigma\right)\frac{dg^\sigma(c_n)}{d\sigma} dc_n +$$
$$\int_{c_L}^{\alpha^*}\left(\frac{d\alpha^*}{d\sigma}\phi_\alpha\left(\alpha^*, c_n, M, \sigma\right) + \phi_\sigma\left(\alpha^*, c_n, M, \sigma\right)\right) g^\sigma(c_n)\, dc_n,$$

The last term is positive due to Assumption 2.(C) (see also the proof of Lemma 8), therefore a sufficient condition for $\frac{dM_S}{d\sigma} > 0$ is that the first two terms are positive. Because $\frac{dc_L}{d\sigma} = \frac{1}{\sigma}(c_L - \bar{c}) < 0$ and $\frac{d}{d\sigma} g^\sigma(c_n) = -\frac{1}{\sigma}\left(g^\sigma(c_n) + (c_n - \bar{c})\, dg^\sigma(c_n)/dc_n\right)$, and remembering that $\phi(\alpha^*, \alpha^*, M, \sigma) = 0$, then under Assumption 2.(D) $\frac{dM_S}{d\sigma} > 0$ is always satisfied.

(Second inequality) If $\Phi > 0$, then all firms stay in the market: $\alpha^* = c_H$ and therefore $\frac{d\alpha^*}{d\sigma} > 0$. Otherwise if $\Phi < 0$, then $\frac{d\alpha^*}{d\sigma} = \frac{d\alpha^*}{dM}\frac{dM_S}{d\sigma} < 0$, from Lemma 7. ■

Proof of Proposition 2. The second and third inequalities are obtained by totally differentiating (19). The first inequality comes from the observation that dividing both sides of (21) the left-hand side is independent of D. ■

Proof of Lemma 9. Profit (23) can be decomposed as:

$$E\left[\tilde{\Pi}_n\right] = \frac{Dt}{N^2} - F +$$
$$\frac{D}{2t}\left\{E\left[\tilde{p}_n\tilde{p}_{\langle n-1\rangle}\right] - 2E\left[\tilde{p}_n^2\right] + 2E\left[\tilde{c}_n\tilde{p}_n\right] - E\left[\tilde{c}_n\tilde{p}_{\langle n-1\rangle}\right] - E\left[\tilde{c}_n\tilde{p}_{\langle n+1\rangle}\right] + E\left[c_n\tilde{p}_{\langle n+1\rangle}\right]\right\}. \tag{40}$$

The expected values in (40) are derived in different parts.

Part 9.I *Under Assumption 1, if firms set their prices following Lemma (2), then* $E\left[(\tilde{p}_n)^2\right] = \sigma^2 W_{I.0}(N) + \bar{c}^2 + 2\bar{c}\hat{k} + \hat{k}^2$ *where* $W_{I.0}(N) = \sum_{i=\underline{l}}^{\bar{l}} w^2_{\bar{l}-|i|} > 0$.

Derivation of Part 9.I. From (**3**): $E\left[(\tilde{p}_n)^2\right] = E\left[\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i\rangle} + \hat{k}\right)^2\right]$ which can be decomposed in three terms: $E\left[\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i\rangle}\right)^2\right] + 2E\left[\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i\rangle}\right)\right]\hat{k} + \hat{k}^2$. Under Assumption **1**, we know that $E\left[\tilde{c}_n^2\right] = \left(\sigma^2 + \bar{c}^2\right)$, and, since $\tilde{c}_n$ and $\tilde{c}_{\langle n+i\rangle}$ are uncorrelated: $E\left[\tilde{c}_n\tilde{c}_{\langle n+i\rangle}\right] = \bar{c}^2$, with $i \in \mathbb{L}_0$. After rewriting $E\left[\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i\rangle}\right)^2\right]$ as:

$$\sum_{i=\underline{l}}^{\bar{l}} \sum_{j=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} w_{\bar{l}-|j|} E\left[\tilde{c}_{\langle n+i\rangle} \cdot \tilde{c}_{\langle n+j\rangle}\right],$$

only two types of terms are left: $E\left[\tilde{c}_n^2\right]$ and $E\left[\tilde{c}_n\tilde{c}_{\langle n+i\rangle}\right]$, $i \in \mathbb{L}_0$. Further, $\sum_{i=\underline{l}}^{\bar{l}} \sum_{j=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} w_{\bar{l}-|j|} = 1$. Since the total weight involving $E\left[\tilde{c}_n^2\right]$ terms is $W_{I.0}(N) = \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|}^2$, we immediately obtain the thesis. $\diamondsuit$

Part 9.II *Under Assumption 1, if firms set their prices following Lemma 2, then* $E\left[\tilde{p}_n\tilde{p}_{n-1}\right] = \sigma^2 W_{I.1}(N) + \bar{c}^2 + 2\bar{c}\hat{k} + \hat{k}^2$ *where* $W_{I.1}(N) = \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} w_{\bar{i}-|i+1|} > 0$.

Derivation of Part 9.II. From (3):

$$E\left[\tilde{p}_n\tilde{p}_{n-1}\right] = E\left[\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i\rangle} + \hat{k}\right)\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{\langle n+i-1\rangle} + \hat{k}\right)\right].$$

The decomposition of the expected value is similar to that presented in the proof of Part **(9.I)**. To obtain the proof of the proposition, note that the total weight involving $E\left[\tilde{c}_n^2\right]$ terms is $W_{I.1}(N) = \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} w_{\bar{l}-|i+1|}$. $\diamondsuit$

Part 9.III *Under Assumption 1, if firms set their prices following Lemma 2, then:* ***a)*** $E\left[\tilde{c}_n\tilde{p}_n\right] = \sigma^2 w_{\bar{l}} + \bar{c}^2 + \bar{c}\hat{k}$*; and* ***b)*** $E\left[\tilde{c}_n\tilde{p}_{\langle n-1\rangle}\right] = \sigma^2 w_{\bar{l}-1} + \bar{c}^2 + \bar{c}\hat{k}$.

Derivation of Part 9.III. **a)** $E\left[\tilde{c}_n\tilde{p}_n\right] = E\left[\tilde{c}_n\left(\sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \tilde{c}_{n+i} + \hat{k}\right)\right]$. The thesis follows from the same reasonings of Parts (9.I) and (9.II) by noting that the weight associated to $E\left[\tilde{c}_n^2\right]$ is $w_{\bar{l}}$.

b). Similar to a). In this case the weight of $E\left[\tilde{c}_n^2\right]$ is $w_{\bar{l}-1}$. $\diamondsuit$

To complete the proof of the Lemma, consider that firms lack information on their own and their opponents' costs and locations; therefore $E\left[\tilde{p}_n\tilde{p}_{\langle n+1\rangle}\right] = E\left[\tilde{p}_n\tilde{p}_{\langle n-1\rangle}\right]$ and $E\left[c_n\tilde{p}_{\langle n+1\rangle}\right] = E\left[c_n\tilde{p}_{\langle n-1\rangle}\right]$. After substituting the results from Parts 9.I, 9.II, and 9.III into (40), most of the terms cancel out, finally yielding: $E\left[\tilde{\Pi}_n\right] = \frac{D}{t}\left(\hat{k}^2 + \sigma^2 W_I(N)\right) - F$, where $W_I(N) := W_{I.0} - W_{I.1} + w_{\bar{l}} - w_{\bar{l}-1} = \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|}\left(\left(w_{\bar{l}} - w_{\bar{l}-|i|}\right) - \left(w_{\bar{l}-1} - w_{\bar{l}-|i+1|}\right)\right)$. To compute $\bar{W}_I$, note that $W_I(N) = \left(w_{\bar{l}} - w_{\bar{l}-1}\right) - \sum_{i=\underline{l}}^{\bar{l}} w_{\bar{l}-|i|} \cdot \left(w_{\bar{l}-|i|} - w_{\bar{l}-|i+1|}\right)$. Using Lemma 1 and the numerical algorithm to calculate the weights described in fn. 7, $z = x^{-1}$, for N large enough $W_I(N) \simeq (y - zy) + \sum_{i=\underline{l}}^{\bar{l}}\left(yz^{|i|} \cdot yz^{|i+1|}\right) - \sum_{i=\underline{l}}^{\bar{i}}\left(yz^{|i|}\right)^2 = (y - zy) + y^2\frac{z-1}{z+1} = 1 - \frac{4}{9}\sqrt{3} \simeq 0.23020$. ■

Proof of Proposition 3. It derives directly from Lemma 9. ■

Proof of Proposition 4. Using (33), (34), (36) and Assumption (1), the expected average transport cost $E[TC]$ and the expected average (variable) cost of production $E[VC]$ are:

$$E[TC] = \frac{t}{4N} + \frac{1}{2}\frac{N}{t}\sigma^2;\ E[VC] = \bar{c} - \frac{N}{t}\sigma^2.$$

Finally, fixed costs per consumer are: $FC = NF/D$. Hence, the expected average social cost:

$$E[C] = E[TC + VC + FC] = \left(\frac{t}{4N} + \frac{1}{2}\frac{N}{t}\sigma^2\right) + \left(\bar{c} - \frac{N}{t}\sigma^2\right) + \frac{NF}{D} \tag{41}$$

■

References

[1] Alderighi, Marco and Piga, Claudio A. 2008. "The circular city with heterogeneous firms. " (July, 13 2008). Available at SSRN: http://ssrn.com/abstract=1159381

[2] Alderighi, Marco and Piga, Claudio A. 2010. "On cost restrictions in spatial competition models with heterogeneous firms."*Economics Letters,* 108(1):40-42

[3] Becker, Gary. S. 1983. "A Theory of Competition among Pressure Groups for Political Influence ". *Quarterly Journal of Economics*, 98: 371–400.

[4] Chamberlin, Edward, H. 1933. "*The Theory of Monopolistic Competition*". Harvard University Press: Cambridge.

[5] Deneckere, Raymond and Rothschild, Michael, 1992. "Monopolistic Competition and Preference Diversity ". *Review of Economic Studies*, 59: 361–373.

[6] Eaton, Curtis B. and Lipsey, Richard G., 1978. "Freedom of Entry and the Existence of Pure Profit". *Economic Journal*, 88(351): 455–69.

[7] European Commission. 2005. "Professional Services - Scope for more reform". *Communication from the Commission*, September, COM. 405.

[8] European Commission. 2004. "Report on Competition in Professional Services". *Communication from the Commission*, February, COM. 83.

[9] Gu, Yiquan and Wenzel, Tobias. 2009. "A note on the excess entry theorem in spatial models with elastic demand." *International Journal of Industrial Organization*, 27: 567–571.

[10] Holmstrom, Bengt, 1999. "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66: 169-82.

[11] Hsieh, Chang-Tai and Moretti, Enrico. 2003. "Can Free Entry be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry". *Journal of Political Economy*, 111: 1076–1122.

[12] Kleiner, Morris M. 2000. "Occupational Licensing". *The Journal of Economic Perspectives*, 14(4):189–202

[13] Leland, Hayne E.. 1979. "Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards". *Journal of Political Economy*, 87: 1328–1346.

[14] Levin, Dan and Peck, James. 2003. “To Grab for the Market or to Bide One’s Time: A Dynamic Model of Entry”. *Rand Journal of Economics*, 34(3): 536–556.

[15] Peltzman, Sam, 1976. “Toward a more General Theory of Regulation ”. *Journal of Law and Economics*, 19: 211–240.

[16] Philipsen, Niels. F. 2010. “Regulation of Liberal Professions and Competition Policy: Developments in the EU and China.” *Journal of Competition Law and Economics*, 6:203–231.

[17] Posner Richard A., 1974. “Theories of Economic Regulation ”. *Bell Journal of Economics*, 5:335–358.

[18] Rothschild, Robert, 1982, “Competitive Behaviour in Chain-Linked Markets”. *Journal of Industrial Economics*, 33(1-2): 57–67.

[19] Salop, Steven C. 1979. “Monopolistic Competition with Outside Goods.” *Bell Journal of Economics*, 10(1): 141–156.

[20] Schaumans, Catherine. 2009. “Strategic Interaction between General Practitioners and Specialists - implications for gatekeeping.”, *mimeo*, Tilburg University.

[21] Schaumans, Catherine and Verboven, Frank. 2008. “Entry and Regulation - Evidence from the Health Care Professions.”, *Rand Journal of Economics*, Winter, 39(4):949–972.

[22] Spulber, Daniel F. 1995. “Bertrand Competition when Rivals’ Costs are Unknown”. *Journal of Industrial Economics*, XLIII(1): 1-11.

[23] Shapiro, Carl. 1986. “Investment, Moral Hazard and Occupational Licensing.” *Review of Economic Studies*, 53: 843–862.

[24] Stigler, George J., 1971. “The Theory of Economic Regulation ”. *Bell Journal of Economics*, 2:3–21.

[25] Syverson, Chad 2004. “Market Structure and Productivity: A Concrete Example”. *Journal of Political Economy*, 112(6):1181–1222.

[26] Tirole, Jean 1988. “*The Theory of Industrial Organization*”. MIT Press: Cambridge, Mass., USA.

[27] Vogel, Jonathan 2008. “Spatial Competition with Heterogeneous Firms”. *Journal of Political Economy*,116(3):423–466.

Tables and Figures

Table 1: The sequence of equilibrium parameters in the Entry Games (percentage values).

N	W_I	W_L
2	22.222	11.111
3	24.000	8.000
4	23.611	6.250
5	23.269	5.540
6	23.111	5.284
7	23.051	5.197
8	23.030	5.169
9	23.023	5.161
10	23.021	5.158
11	23.020	5.157
12	23.020	5.157
13	23.020	5.157
14	23.020	5.157
15	23.020	5.157
20	23.020	5.157
∞	23.020	5.157

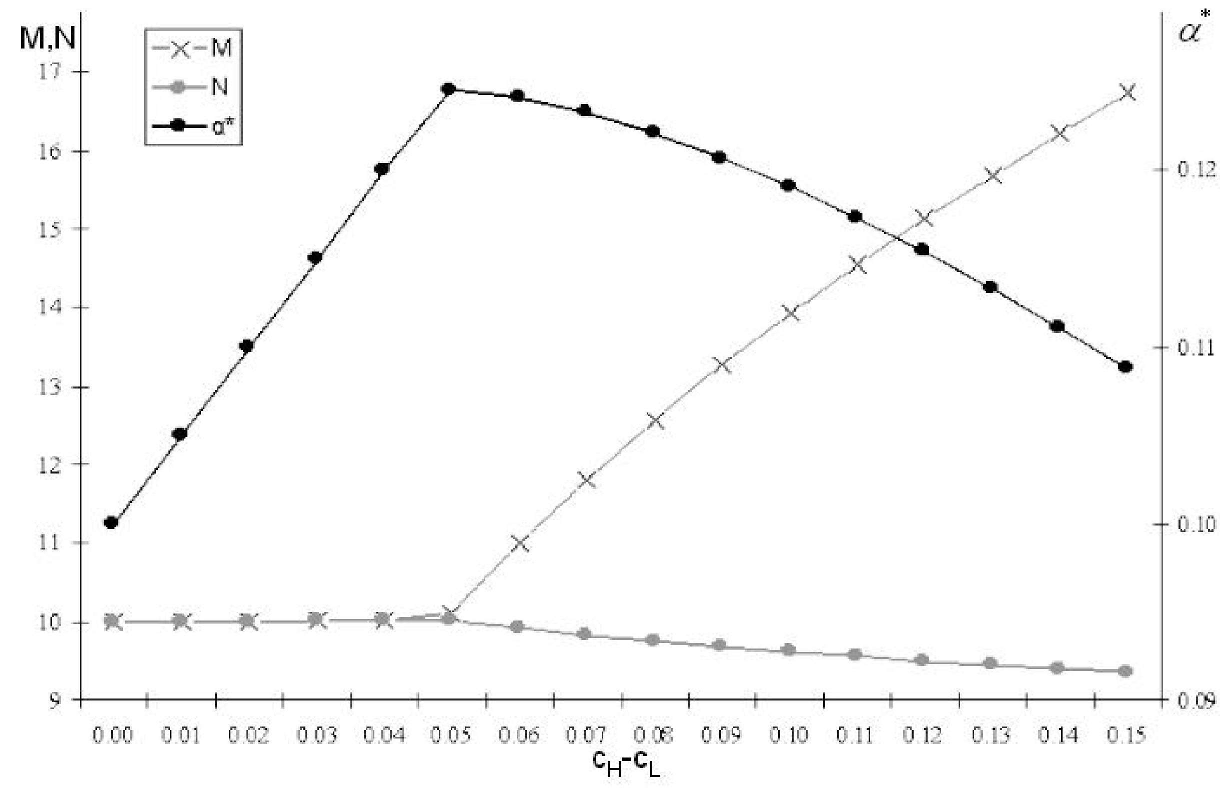

Figure 1: The optimal number firms in the pre-entry stage M and in the production stage N (left scale). Optimal entry threshold (right scale).

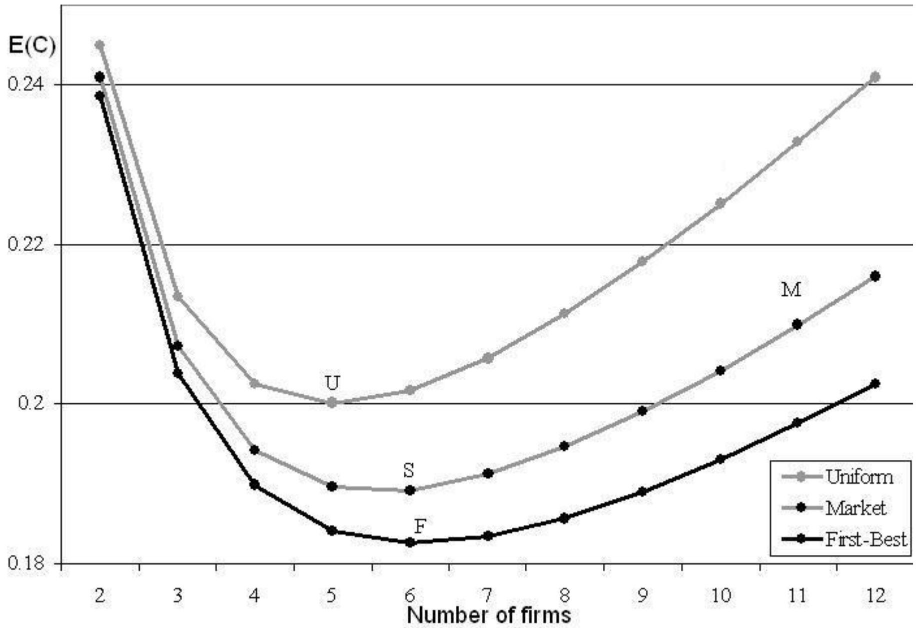

Figure 2: Expected average total cost under different pricing scenarios. F is the social optimum; M is the free-market outcome; S is the second best solution (the social planner's choice of entrants when they set market prices). Under uniform pricing, the planner's choice is in U .