



HAL
open science

Entry and Submarket Concentration: Empirical Evidence from the Pharmaceutical Industry

Maria Letizia Giorgetti

► **To cite this version:**

Maria Letizia Giorgetti. Entry and Submarket Concentration: Empirical Evidence from the Pharmaceutical Industry. *European Review of Industrial Economics and Policy* , 2013, 6. hal-03470102

HAL Id: hal-03470102

<https://hal.science/hal-03470102v1>

Submitted on 8 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Entry and Submarket Concentration: Empirical Evidence from the Pharmaceutical Industry

Maria Letizia Giorgetti

Università degli Studi di Milano; Università Cattolica del Sacro Cuore di Milano; letizia.giorgetti@unimi.it.

The paper investigates the decisions of companies to enter new submarkets, analyzing the relationship between submarket concentration and entry through the explanatory power of submarkets. The relationship is analyzed both at the aggregate level for all countries and at the country-specific level using a pharmaceutical company dataset gathered in seven countries in the 1987-1998 period. A Bayesian panel probit model enables us to measure entry into a new submarket using the lagged dependent variable to incorporate potential state persistence. The results obtained at the submarket level indicate a negative correlation between concentration and entry by diversification, although previous literature at market level has indicated conflicting results for the same relationship.

Entry, Diversification, Bayesian Inference, Panel Data, Probit Model, Random Effects

Ingresso, Diversificazione, Inferenza Bayesiana, Dati Panel, Modello Probit, Effetti Random

The author would like to thank Gianni Amisano for the helpful discussion and for sharing his codes. I am indebted to Michele Grillo and two anonymous referees for their very useful suggestions. I would also like to thank Laura Rondi and the participants of the *IIOC (International Industrial Organization Conference)* held in Chicago on 23-24 April 2004, particularly Paul Walsh and Franco Mariuzzo for their insightful comments. The author also thanks the participants of the seminars held at the Bocconi, Brescia, Pisa, Padua and Milan universities, Joseph A. Clougherty for his detailed discussion at *IIOC 2005*, Atlanta, 8-9 April 2005, and the comments received at the *SIE (Economics Italian Society)*, Naples 21-22 October 2005. A special mention goes to Fabio Pammolli, Director of the Epris Project during which the dataset was collected and organized.

Introduction

The paper aims to provide empirical evidence on the negative relationship between entry by diversification and submarket concentration in the pharmaceutical sector.

The analysis adopts a dynamic Bayesian panel probit approach to model the conditional probability of entering at least one new submarket. The model control for firm features and the covariates incorporate the lagged entry decisions in order to capture potential state persistence.¹

The industrial organization (IO) literature has investigated the relationship between entry and market concentration both theoretically and empirically. However, the concrete empirical relationship is not as easy to establish as it seems. What is the effect of market

¹ Including state persistence aims to measure the tendency of companies that entered a new submarket in the immediate past to enter new submarkets in the successive period.

concentration on entry? Does greater market concentration imply higher markups over marginal costs, driving up profits and thus attracting new entrants? Moreover, can market concentration act as an entry barrier to the extent that it reflects a first-mover advantage of incumbents or superior product differentiation?

The relationship between entry and market concentration was investigated in the 1960s by structure-conduct-performance researchers, whose conclusions often pointed to a negative relationship between market concentration and entry. That conclusion was reached successively by Barbosa (2003), who also adopted a structural approach. Barbosa observes that, in the presence of high market price-cost margins and high market concentration, the incumbents have a greater propensity to engage in strategic entry behavior. “Conversely, low industry concentration and/or low incumbents profitability are likely to discourage incumbents from engaging in strategic behavior, since expected benefits from that may not cover its costs” (Barbosa, 2003).

The role of market concentration can change according to the type of entry. Mueller (1991)² outlined three types of entry: entry by acquisition, greenfield entry, and entry by diversification. A “greenfield entry”,³ whereby a company taps into an unexploited market, will be most unlikely in highly concentrated markets, while entry by acquisition tends to correlate positively with a high level of market concentration (Baldwin, 1995). On the other hand, in terms of entry by diversification, high concentration usually slows entry rates although the presence of other factors can determine the opposite effect. In fact, according to some authors, diversification strategies are more likely in industries (markets) of high growth, high profits (Berry, 1975; MacDonald, 1984), and high concentration. MacDonald (1984) found that diversification into an industry is associated “with declining concentration in industries of initially high concentration, and it increases in industries of initially low concentration”. The literature provides ambiguous results on the role of market concentration in the case of entry by diversification.

To enrich this line of research, it is necessary to shift the focus from market level to “submarket” level, therefore, in this paper entry by diversification is defined as the company’s decision to enter one or more new submarkets. The aim is to assess whether the submarket power of incumbents attracts or discourages competitors.

Naturally, some authors may deem the term “submarket” not sufficiently precise or standard in the literature, mostly because many papers discuss entry in many markets, such as that by Jia (2008) on entry by discount stores or the working paper by Tom Holmes (Holmes, 2011) on dynamic multi-market entry, the entry paper by Ellickson, Houghton and Timmins (2007), and the paper by Aguirregabiria and Mira (2007), all of which study multi-market entry and disregard the term submarkets. However, the distinction between markets and submarkets cannot be interpreted simply as a boring academic debate on terminology.

According to Sutton (1998), the submarkets acquire a primary role in explaining global market structure and evolution if related to sectors with high R&D intensity. So what exactly are “submarkets”?

² Mueller (1991) enumerates two other types of entry: spin-off and entry through a foreign company; in this paper the three types enumerated have priority.

³ See Amisano and Giorgetti (2013a) on the different meanings of greenfield entry.

The standard theory and the antitrust authorities adopt a definition of market as comprising a set of goods, all of which are substitutes. Thus, submarkets can be defined as market segments where products are connected on the demand side.

As Sutton (Sutton, 1998) pointed out “once one turns to the case of R&D intensive industries, the formulation of the theory developed above becomes inadequate. In order to have an adequate picture of what really happens in an industry technological trajectories and their associated submarkets should be the right level of study” (Sutton, 1998). The idea to identify submarkets by taking into account solely the demand side may overlook the presence of technical trajectories belonging to different groups of products. The “submarkets framework” (Sutton, 1998) implies that entry and industry concentration are better understood if the focus is extended from different markets to incorporate different submarkets.

The theoretical approach can be briefly explained as follows: according to Sutton (1998, p. 63), the parameter α sets a lower limit for the CR1, that is, the rate of the total market sales generated by the leading company in the market. The global industry CR1 depends on the linkages among submarkets on the demand and supply sides. On the demand side, the linkages among submarkets depend on the willingness of consumers to pay for an increase in technical performance and on the level of consumption-substitution among goods associated with different technical trajectories; on the other hand, the supply side, linkages among submarkets depend on the presence of scope economies in R&D among different trajectories⁴ and on the way the R&D investments impact technical performance. Given the intensity of R&D, different linkages among submarkets impact the global level of CR1 (at industry level): a lower CR1 is the result of weak submarket linkages while a higher CR1 is the result of strong submarket linkages.

That clarifies the important role of submarkets in shaping the level of concentration of an entire industry. The analysis of entry by diversification in high R&D investment sectors, such as the pharmaceutical sector, helps to shed light on the actual linkages among submarkets.

Which brings us to the key empirical question: what role does submarket concentration play on entry by diversification when moving from market-level concentration to submarket-level concentration? Ex-ante, the effect of submarket concentration on entry can go either way: it could either signal the appeal of different classes of drugs, leading high concentration to spur entry decisions that dynamically enhance entry and, therefore, diversification at the firm level; or high submarket concentrations could be the result of high entry barriers in each submarket, with submarket-level concentration acting as a potential deterrent to entry decisions and therefore diversification at the firm level.

It is particularly important to investigate the role of the incumbents, (in other words, the impact of submarket power), given the recent transformations in the pharmaceutical sector’s R&D process, which seem to have eliminated durable long-term first-mover advantages exportable to different submarkets (Sutton, 1998).

The paper thus investigates whether the role played by concentration in different submarkets has a significant and consistent impact on entry into at least one new submarket, (intended as entry by diversification) regardless of the submarket considered.

⁴ See Sutton (1998): the author illustrates the presence of low levels of scope economies in R&D among submarkets in the chapter on the pharmaceutical sector. This point is discussed in the literature, while conflicting conclusions are drawn by Cockburn and Henderson in their paper of 1996 (Cockburn and Henderson, 1996). See also Giorgetti (2006).

The paper is organized as follows: section one addresses the methodological issues involved in an analysis of this type; section two reviews the specific literature; section three provides an overview of the pharmaceutical market and defines the dataset used in the application; section four briefly presents the econometric model; section five discusses the results and the theoretical implications; and the last section presents the conclusions and implications of the analysis.

1. Methodological hurdles

The analysis suffers from different methodological problems of both the economic and the econometric type. At the submarket level of the analysis, moving from theory to applied research requires filling the gap of submarket identification. Regardless of the theory adopted, applied researchers face data-availability hurdles and the need to rely on official classifications. As in earlier work, the paper again adopts the one digit ATC (anatomical therapeutical classification) classification (see section three and the appendix), although the adoption of one digit classification is far from ideal because the submarket definition remains too broad. Future research will be based on a finer three digit segmentation. However, the identification of one digit ATC classification does not invalidate the presence of different R&D trajectories for products belonging to different submarkets. For example, looking at the two submarkets (see the appendix): G “genito urinary system and sex hormones product” and H “systemic hormonal preparation (excluding sex hormone) products” could imply the existence of a single submarket designated “hormone products”, so the entry in submarket H could be not entry by diversification into a new submarket with a different technical trajectory, but it could be merely a matter of product differentiation. However, that is not strictly true because the systemic hormonal preparation (submarket H), as suggested by the word “systemic”, has an impact on the entire human body and does not act at the localized level like the products belonging to genito urinary system and sex hormone products do.

From an econometric viewpoint, the literature (see also Amisano and Giorgetti, 2008, 2013a, 2013b) shows that this field of research does not adopt the predominant structural approach in IO, but attempts to overlap the potential limits of reduced-form models with a Bayesian approach where an attempt is made to extract information from the data. While advocates of the structural approach might criticize the lack of massive economic theory foundations, the Bayesian approach definitely reduces the distance between the data and the estimated model. Another potential econometric problem could derive from how the estimation has been performed. As mentioned earlier, the dataset is made up of information on the sales generated by individual companies in different submarkets and different countries (Germany, Italy, UK, France, Canada, Spain, and the United States) in the 1987-1998 period. The estimates are based on the separate use of the aggregate dataset (where company sales are aggregated by country for each submarket in each year) and each single country dataset. This could attract criticism due to the exclusion of country fixed effects, but is offset by the country-specific as well as the aggregate evidence provided. Clearly, incorporating country fixed effects would make the empirical part more compact, but there are two good reasons to use separately the aggregate dataset and evidence for each country. First, the paper purposefully presents the analysis for each country separately,⁵ and, second,

⁵ The decision to take the single country approach ensues from the fact that drug regulations change from country to country over time. The literature has, in fact, emphasized the impact of different price regulations

more importantly, there is a theoretical reason. According to Sutton's framework (1998), when analyzing sectors with high R&D investments, given the worldwide level of spillovers, it is important to perform the analysis by considering each industry at worldwide level as a single market. Thus, even without data pertaining to Japan and Switzerland, the seven countries in the dataset represent an important portion of the global scenario as they cover the largest key pharmaceutical markets.

2. Literature review

A review of the literature has the aim of providing an overview of earlier contributions on the subject. The intuition that an industry consists of differentiated submarkets is not new but some of the literature emphasizes the relevance of submarkets in assessing industry regularities, including: entry, exit, diversification and firm growth (Klepper and Thompson, 2006).

All these regularities had been investigated already by the IO literature but, successively, Sutton (1998) and Klepper and Thompson (2006) emphasized the explanatory power of submarkets in spelling out how market structure and evolution work.

Given that the literature does not sufficiently emphasize the "submarket approach", using Sutton's "submarkets framework" enables us to reinterpret market structure regularities by adopting a synergic fusion of different strands of IO literature: the structure-conduct-performance approach that adopts cross-section analysis, the mathematics of game theory, and Dosi's (1982) contribution of technological trajectories. Since the 1970s, the mathematics of game theory has been used only in intra-industry case studies due to the impossibility of using the same models for inter-industry studies. Sutton (1998) suggests, adopting game theory, to recover inter-industries regularities. The regularities across industries are few but the correct objective is to verify the presence of this small number of regularities across industries. Moreover, Sutton's approach encompasses the important role of history in shaping market structure and its evolution, emphasizing the important contribution to IO highlighted by evolutionary economics (Dosi, 1982).

This paper briefly presents the results on entry by diversification that can be obtained by adopting the submarkets as a key of investigation. The aim is to emphasize how the focus on submarkets, especially submarket concentration, augments our understanding of the impact of concentration on entry by diversification.

There are several papers on entry generally and a good number of papers on the pharmaceutical industry, these include: Scott Morton (1999), who examines the introduction of a generic drug in the U.S. market; Kyle (2006), who studies the determinants of new drug launches in the largest drug markets of the G7 countries; Lanjouw (2005), who investigates how patent rights and price regulation influence both the introduction and how quickly new drugs are marketed in a country; and Danzon, Wang and Wang (2005), who study the impact of price regulation on the launch timing of a new drug. The number of papers on the pharmaceutical sector decreases sharply in terms of contributions that address the explanatory power of submarkets, although we point out that of Bottazzi *et al.* (2001).

on entry decisions. It seems that regulatory policies have become increasingly important in driving the evolution of the pharmaceutical market structure, therefore, to control for different regulatory environments it was necessary to conduct the analysis also at the single country level.

In terms of the submarket approach, earlier papers by Amisano and Giorgetti (2008, 2013a, 2013b) adopt a reduced form approach while recent applied IO literature is taking an increasing interest in the structural approach framework.

The *Handbook of Econometrics* publishes papers by Akerberg *et al.* (2007) who present a very exhaustive discussion of entry in dynamic models in different industries adopting a structural approach; Panle (2008), who addresses entry by discount stores; Holmes (2011), who analyses dynamic multi-market entry; and entry papers by Aguirregabiria and Mira (2007) and Ellickson, Houghton and Timmins (2007).

The work by Reiss and Wolak (2007) in the *Handbook of Econometrics* underscores the importance of combining economic theory and econometric techniques to answer empirical questions in IO. This can be done, according to these authors, by adopting a structural approach, although the approach cannot solve other kinds of problems, e.g., the difficulty of translating all the economic theories into estimable relationships, the lack of data for some specific variables in an economic theory, and, lastly, the modeler's discretion when it comes to including variables not explicitly part of the theory or on the errors term. This author's field of research does not adopt the predominant structural approach but does provide a useful contribution to the entry process, and answers the following question: what role does submarket concentration play in spurring or deterring entry decisions into (at least) one new submarket?

The main results of the paper indicate that submarket concentration has a significant discouraging effect on the decision to enter new submarkets.

That result is fairly unsettling, given that the aggregate dataset and all single country datasets indicate that submarket concentration acts as a barrier to entry, the same role discussed by the literature in terms of market concentration generally, without differentiating between entry types.

The average partial effect of this variable is noticeable and determines relevant negative elasticities of entry probability with respect to existing submarket concentration: the effect of the high entry barriers seems to outweigh the attractiveness of existing market power.

3. The pharmaceutical sector and the dataset

The paper provides an overview of the pharmaceutical sector as background for the analysis. The pharmaceutical sector is a science-based industry characterized by high investments in R&D (Matraves, 1999; Sutton, 1998). The sector is characterized by an oligopolist core that undertakes innovative R&D and by a second group of companies that undertake "imitative" R&D. The two groups of companies, in and outside the core, are not evenly distributed across countries; while the first group includes companies primarily from the U.S., the UK, and Germany, the second group contains companies from France and Italy (Di Masi, 2001).

Because of different types of research (pioneering or imitative R&D) and different drug regulations across countries, the analysis model is based on each of the seven countries taken individually.

A detailed sales dataset for all the most important international companies in different submarkets and different countries is used for the 1987-1998 period. The data was obtained under the aegis of a research project, part of the EPRIS project of the University of Siena.⁶

⁶ The data was collected by IMS Health, the leading worldwide provider of information solutions to the pharmaceutical and health care industries.

The global dataset contains annual data on the sales of two hundred and eight international companies operating in sixteen “submarkets” and in seven countries: Canada, France, Germany, Italy, Spain, UK and the U.S.⁷ The submarkets are identified according to the ATC classification. This classification is based on the EphMRA (European Pharmaceutical Marketing Research Association) anatomical classification. In particular, first level segmentation, where products are grouped according to the principal anatomical group (see the appendix), is used.

The “global” dataset (where all the countries together are considered as single geographical entities) was obtained by merging the seven different datasets into one.

Each national dataset contains information on the sales of all the international companies operating in that particular country in the 1987-1998 period. The number of international companies varies by country: Canada has seventy-one, France seventy-five, Germany seventy-three, Italy seventy-one, Spain eighty-one, the UK seventy, and the U.S. fifty-seven.

Returning to the dynamics of entry decisions, it is necessary to build the variables starting with the information available in the dataset.

The company specific entry variable (*ENTRY*) is a discrete variable y_{it} which is equal to one if and only if company i at time t has a positive sales volume in at least one submarket that produced zero sales the year before.

According to this definition of entry, table 1a presents data calculated on the number of submarkets in which companies are active (i.e., they have a non-zero volume of sales) at the aggregate multi-country level. The table indicates for each year the median value of the submarkets entered by each company in the previous year for all seven countries together. This is a very succinct measure of the achieved diversification level. The median value ranges from eight submarkets in 1988 to nine submarkets in 1998, suggesting the existence of an “optimal” level of diversification (Amisano and Giorgetti, 2013b).

⁷ For each year, the sales of each international company in each principal anatomical group were obtained by summing the sales of all the national subsidiary companies belonging to it. For each national company, the international parent company is known and therefore aggregation across different countries is straightforward. Sales is the only variable contained in the dataset. For computational reasons, the one digit ATC classification is used. For each country all sales are expressed in U.S. dollars to ensure homogeneity. Data are transformed in real terms by using the U.S. GDP deflator.

Table 1.a. Number of lagged submarkets in which companies are active, all countries dataset

Year	Mean	Modal	Median
1988	8.23	7	8
		12	9
1994	8.61	9	9
1995	8.65	9	9
1996	8.74	9	9
1997	8.81	9	9
1998	8.85	9	9

Source: My elaboration on IMS Health data

Table 1b shows the transition probabilities during the review period for the variable *ENTRY* for the global dataset: when a company makes a succession of entries year after year, the lower left cell of the matrix is incremented by one. An initial observation is that the unconditional probability of entry is quite low (13.7%) and that the probability of entry seems to be positively influenced by the entry made the previous year: the probability of entry increases from 12.9% to 18.5% when entry is controlled for being equal to one, instead of zero, in the previous period.⁸ That implies that companies that have recently entered a new submarket are slightly more likely to continue doing so, although the econometric analysis performed does not confirm this aspect when controlled for other covariates.

Table 1.b. Transition Probabilities, entry decisions

Entry		0	1
	0	87.11	12.89
	1	81.47	18.53
uncond. prob		86.30	13.70
LR test		6.29	
p-value		0.01	

Source: My elaboration on IMS Health data

Table 2 provides a more detailed picture of entry decision descriptions and the relative data on the number of submarkets entered each year from 1988-1998. As usual, the table indicates both multi-country and single-country evidence, the maximum, mean and standard deviation in the number of submarkets entered, and the percentage of companies entering one, two or more submarkets.

At the aggregate level, table 2 shows that the mean value of submarkets entered is often, but not always, around one. Mean values of around one are found also in the single country dataset (see tables 2 to 2i), with the notable exception of 1991 in Germany, with a mean of over eight.

⁸ A simple test of the significance of this difference is provided by running a likelihood ratio independence test on the transition probability matrix, which leads to rejection of the independence null at the 5% level, while at the 1% level, the null would be accepted.

Table 2. observed entry decisions

Table 2.a. All countries	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	4	8	5	3	2	2	15	4	2	2	2
Mean	1.33	2.44	1.41	4.27	0.98	1.03	6.96	1.30	1.02	1.18	1.17
Standard Dev	0.48	0.71	0.52	1.07	0.38	0.42	1.08	0.49	0.43	0.42	0.30
1 new submarket %	13.94	11.54	14.90	12.50	12.98	13.94	7.69	13.46	14.42	10.10	5.77
2 new submarkets %	0.00	1.92	0.96	3.85	0.96	1.44	0.96	1.44	1.92	2.40	0.96
Table 2.b: Canada	1988	1989	1990	1991	1992	1993 1989	1994 1991	1995 1993	1996	1997 1997	1998
Maximum	1	2	5	2	1	2	3	4	1	2	2
Mean	0.90	0.95	2.01	1.20	0.87	1.00	1.36	1.66	0.87	1.03	1.04
Standard Dev	0.32	0.43	0.73	0.50	0.35	0.41	0.48	0.78	0.35	0.48	0.40
1 new submarket %	11.27	16.90	14.08	12.68	14.08	14.08	12.68	15.49	14.08	16.90	12.68
2 new submarkets %	0.00	1.41	2.82	4.23	0.00	1.41	0.00	5.63	0.00	2.82	1.41
Table 2.c: France	1988	1989	1990	1991	1992	1993 1989	1994 1991	1995 1993	1996	1997 1997	1998
Maximum	2	7	1	2	2	2	2	1	1	1	1
Mean	0.88	3.63	0.90	1.19	1.01	1.22	1.08	0.87	0.90	0.92	0.99
Standard Dev	0.46	0.85	0.31	0.33	0.40	0.34	0.38	0.36	0.33	0.27	0.20
1 new submarket %	21.33	10.67	10.67	6.67	13.33	6.67	10.67	14.67	12.00	8.00	4.00
2 new submarkets %	1.33	0.00	0.00	1.33	1.33	1.33	1.33	0.00	0.00	0.00	0.00
>2 new submarkets %	0.00	1.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Table 2.d: Germany	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	3	3	2	13	2	2	14	1	2	1	2
Mean	1.41	1.57	0.99	8.03	1.25	1.18	8.68	0.93	1.14	0.93	1.23
Standard Dev	0.46	0.44	0.42	2.17	0.32	0.44	1.65	0.30	0.35	0.30	0.43
1 new submarket %	10.96	8.22	15.07	17.81	5.48	10.96	12.33	9.59	8.22	9.59	9.59
2 new submarkets %	0.00	0.00	1.37	0.00	1.37	2.74	0.00	0.00	1.37	0.00	2.74
>2 new submarkets %	1.37	1.37	0.00	4.11	0.00	0.00	1.37	0.00	0.00	0.00	0.00
Table 2.e: Italy	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	1	2	2	2	1	1	1	1	2	1	1
Mean	0.85	1.23	0.97	1.04	0.85	0.79	0.95	0.91	0.97	0.91	0.91
Standard Dev	0.36	0.49	0.42	0.40	0.38	0.42	0.26	0.32	0.49	0.30	0.30
1 new submarket %	15.49	11.27	15.49	12.68	16.90	22.54	7.04	11.27	19.72	9.86	9.86
2 new submarkets %	0.00	4.23	1.41	1.41	0.00	0.00	0.00	0.00	2.82	0.00	0.00
>2 new submarkets %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Table 2.f: Spain	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	3	1	1	1	5	1	2	2	1	1	1
Mean	1.43	0.91	0.91	0.88	2.43	0.91	1.67	1.29	0.93	0.89	0.91
Standard Dev	0.44	0.30	0.30	0.33	0.65	0.32	0.25	0.31	0.26	0.35	0.30
1 new submarket %	9.88	9.88	9.88	12.35	8.64	11.11	1.23	4.94	7.41	13.58	9.88
2 new submarkets %	0.00	0.00	0.00	0.00	1.23	0.00	1.23	1.23	0.00	0.00	0.00
>2 new submarkets %	1.23	0.00	0.00	0.00	1.23	0.00	0.00	0.00	0.00	0.00	0.00
Table 2.g: UK	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	3	1	1	1	5	1	2	2	1	1	1
Mean	1.43	0.91	0.91	0.88	2.43	0.91	1.67	1.29	0.93	0.89	0.91

Standard Dev	0.44	0.30	0.30	0.33	0.65	0.32	0.25	0.31	0.26	0.35	0.30
1 new submarket %	9.88	9.88	9.88	12.35	8.64	11.11	1.23	4.94	7.41	13.58	9.88
2 new submarkets %	0.00	0.00	0.00	0.00	1.23	0.00	1.23	1.23	0.00	0.00	0.00
>2 new submarkets %	1.23	0.00	0.00	0.00	1.23	0.00	0.00	0.00	0.00	0.00	0.00
Table 2.h: USA	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
Maximum	2	2	3	3	2	2	2	2	3	2	2
Mean	0.99	0.95	1.22	1.57	1.22	1.16	0.96	0.86	1.17	0.89	0.97
Standard Dev	0.62	0.62	0.64	0.55	0.64	0.55	0.58	0.49	0.67	0.54	0.52
1 new submarket %	24.54	26.32	21.05	10.53	15.79	15.79	24.56	24.56	22.81	26.32	21.05
2 new submarkets %	7.02	7.02	3.51	1.75	8.77	5.26	5.26	1.75	5.26	3.51	3.51
>2 new submarkets %	0.00	0.00	1.75	1.75	0.00	0.00	0.00	0.00	1.75	0.00	0.00

As for the percentage of companies entering one new submarket, at the aggregate level (table 2a) the entry flow into one new submarket is quite steady year after year –around 13-14% annually– albeit declining in the most recent years of the sample by around 10% and 6% for 1997 and 1998, respectively. This pattern is not reflected at the single-country level, where the pattern is less stable over time and shows substantial cross-country heterogeneity. In particular, the table shows that very high percentages of companies enter one new submarket each year in the US (see Table 2h), mostly well above 20%.

This seems to be a clear sign of more aggressive entry decisions being implemented in the U.S. Such high values cannot be observed in other countries, such as the UK and Germany, despite the high levels of R&D spending of their pharmaceutical companies.

Returning to our dataset, the variables adopted in the analysis are indicated below; all the regressors are lagged into the model and therefore the first observation (1987) is lost. The variables are:

1. SA_{t-1} (*lagsales*), a measure of size, namely the lagged (at $t - 1$) value of the sales of each company in all the submarkets (zero sales signifies the company has no presence);
2. SU_{t-1} (*lagsub*), the lagged number of submarkets in which each company is present. This is a measure of the diversification already achieved;
3. Y_{it-1} (*lagentry*), the lagged dependent entry variable;
4. $HHI_{i,t-1}$ (*lagHHI*), the lagged HHI (Herfindahl index) (the concentration measure used) that each company faces when it enters the submarket at time t .

In particular, according to the HHI:

1. for a company that entered at least one submarket in the previous period, HHI is averaged over the submarkets the company entered;
2. for a company that did not enter a submarket, HHI is “the maximum at time $t - 1$ across all the submarkets” that the company has not entered. The HHI is computed asymmetrically, with an upward bias if there is no entry. This could mean that the negative estimated relationship between entry and concentration could be distorted by this asymmetric computation.⁹

However, this concern is not justified as a previous analysis¹⁰ has shown that the negative relationship remains even when the minimum HHI is used. The choice of maximum HHI means that submarket concentration is expected to act as a barrier to entry.

When entry fails to occur, the highest submarket concentration across all the submarkets at time $t - 1$ is identified. Nevertheless, further investigation is required on this issue, with future research modeling entry decisions directly at the submarket level. The HHI could be calculated as $HHI_{i,m,t-1}$ where m identifies each specific submarket, eliminating all

⁹ I would like to thank an anonymous referee for pointing out this potential problem.

¹⁰ See Amisano and Giorgetti (2008).

potential construction hurdles. The paper now provides summary evidence on these variables for the dataset used in table 3.

The aggregate evidence is broken down by year and by observations corresponding to entry decisions ($y_{it} = 1$) and observations for which no entry takes place at time t ($y_{it} = 0$). The panel shown in table 3, which summarizes the empirical evidence for the whole of the review period, indicates that *lagsales* and *lagsub* have higher mean values for the $y_{it} = 0$ entries, while the mean value of y_{it-1} (*lagentry*) is lower for the same observations.

This could suggest that size and achieved diversification, as measured by *lagsub*, have a negative effect on the probability of entry, while entry in the previous period positively influenced the entry probability. Nevertheless, entry is eminently a dynamic process and it is therefore useful to consider also the data disaggregated for each year (see table 3). The data follows the same pattern each year with some exceptions: 1992 for *lagsub* and 1998 for *lagsales*. Differences among the groups of observations are not particularly significant, implying that these variables do not play a clear-cut role in a dynamic discrete choice framework. In addition, it suggests the idea of an “optimal” level of diversification, which in turn can be interpreted in light of the notion of an “optimal core” of submarkets to enter.¹¹

In terms of submarket concentration (*lagHHI*), a lower average concentration level for the observations with $y_{it} = 1$ can be observed: this large difference confirms the hypothesis that high concentration acts as an entry barrier, an effect not only confirmed, but also precisely quantified in its quantitative extent by the econometric analysis.

¹¹ More details on these issues can be found in see Amisano and Giorgetti (2013b).

Table 3. descriptive statistics of the regressors, all countries, year by year

		all observations				Y _{it} =1				Y _{it} =0						
Year	Variable	Obs	Mean	S.D.	Min	Max	Obs	Mean	S.D.	Min	Max	Obs	Mean	S.D.	Min	Max
1988	lagsales	208	5.25	2.28	0.00	9.93	26	5.13	1.99	0.47	9.15	182	5.27	2.32	0.00	9.93
	lagsub	208	8.20	4.55	0.00	16.00	26	7.46	3.58	1.00	13.00	182	8.30	4.67	0.00	16.00
	lagentry	208	0.03	0.18	0.00	1.00	26	0.08	0.27	0.00	1.00	182	0.03	0.16	0.00	1.00
	lagHHI	208	0.37	0.1	0.04	0.41	26	0.12	0.1	0.1	0.41	182	0.41	0	0.06	0.41
1989	lagsales	208	5.33	2.21	0.00	9.92	27	4.43	1.90	0.00	7.98	181	5.47	2.22	0.00	9.92
	lagsub	208	8.30	4.52	0.00	16.00	27	6.78	3.42	0.00	13.00	181	8.52	4.63	0.00	16.00
	lagentry	208	0.13	0.34	0.00	1.00	27	0.26	0.45	0.00	1.00	181	0.11	0.31	0.00	1.00
	lagHHI	208	0.35	0.1	0.05	0.39	27	0.12	0.1	0.1	0.39	181	0.38	0	0.06	0.39
1990	lagsales	208	5.33	2.27	-3.81	9.97	31	4.84	2.00	0.00	8.17	177	5.42	2.31	-3.81	9.97
	lagsub	208	8.32	4.45	0.00	16.00	31	7.94	3.90	0.00	15.00	177	8.38	4.54	0.00	16.00
	lagentry	208	0.12	0.33	0.00	1.00	31	0.23	0.43	0.00	1.00	177	0.10	0.30	0.00	1.00
	lagHHI	208	0.32	0.1	0.05	0.36	31	0.12	0.1	0.1	0.36	177	0.35	0	0.05	0.36
1991	lagsales	208	5.38	2.20	-0.98	9.95	31	5.17	2.07	0.00	9.75	177	5.42	2.22	-0.98	9.95
	lagsub	208	8.40	4.42	0.00	16.00	31	8.26	3.95	0.00	14.00	177	8.42	4.50	0.00	16.00
	lagentry	208	0.14	0.35	0.00	1.00	31	0.16	0.37	0.00	1.00	177	0.14	0.35	0.00	1.00
	lagHHI	208	0.29	0.09	0.00	0.32	31	0.09	0.06	0.00	0.32	177	0.32	0.03	0.06	0.32
1992	lagsales	208	5.49	2.13	-0.66	10.08	25	5.31	1.96	0.00	8.89	183	5.52	2.16	-0.66	10.08
	lagsub	208	8.57	4.43	0.00	16.00	25	8.84	4.24	0.00	15.00	183	8.53	4.46	0.00	16.00
	lagentry	208	0.16	0.37	0.00	1.00	25	0.24	0.44	0.00	1.00	183	0.15	0.36	0.00	1.00
	lagHHI	208	0.3	0.1	0.06	0.32	25	0.17	0.1	0.1	0.32	183	0.32	0	0.07	0.32
1993	lagsales	208	5.56	2.17	-2.77	10.22	30	5.43	1.62	2.35	8.92	178	5.59	2.26	-2.77	10.22
	lagsub	208	8.53	4.43	0.00	16.00	30	7.97	3.00	2.00	13.00	178	8.62	4.62	0.00	16.00
	lagentry	208	0.10	0.30	0.00	1.00	30	0.07	0.25	0.00	1.00	178	0.11	0.31	0.00	1.00
	lagHHI	208	0.32	0.1	0.05	0.37	30	0.09	0.1	0.1	0.38	178	0.33	0	0.07	0.33
1994	lagsales	208	5.60	2.12	-0.89	10.26	19	4.84	2.40	0.00	9.85	189	5.68	2.08	-0.89	10.26
	lagsub	208	8.58	4.41	0.00	16.00	19	7.63	4.89	0.00	14.00	189	8.68	4.36	0.00	16.00
	lagentry	208	0.14	0.35	0.00	1.00	19	0.16	0.37	0.00	1.00	189	0.14	0.35	0.00	1.00
	lagHHI	208	0.33	0.1	0.05	0.35	19	0.15	0.1	0.1	0.35	189	0.35	0	0.05	0.35
1995	lagsales	208	5.71	1.98	0.00	10.30	31	5.13	1.66	2.32	9.68	177	5.81	2.02	0.00	10.30
	lagsub	208	8.62	4.34	0.00	16.00	31	7.42	3.62	1.00	14.00	177	8.83	4.43	0.00	16.00
	lagentry	208	0.10	0.30	0.00	1.00	31	0.10	0.30	0.00	1.00	177	0.10	0.30	0.00	1.00
	lagHHI	208	0.3	0.1	0.05	0.34	31	0.14	0.1	0.1	0.34	177	0.33	0.4	0.06	0.34
1996	lagsales	208	5.78	1.97	0.00	10.32	30	5.18	1.94	0.00	10.00	178	5.88	1.96	0.00	10.32
	lagsub	208	8.71	4.29	0.00	16.00	30	7.20	4.26	0.00	15.00	178	8.96	4.25	0.00	16.00
	lagentry	208	0.14	0.35	0.00	1.00	30	0.20	0.41	0.00	1.00	178	0.13	0.34	0.00	1.00
	lagHHI	208	0.3	0.1	0.05	0.33	30	0.12	0.1	0.1	0.33	178	0.33	0	0.07	0.33
1997	lagsales	208	5.89	1.89	0.00	10.33	24	4.79	2.03	0.00	7.81	184	6.03	1.83	2.46	10.33
	lagsub	208	8.78	4.26	0.00	16.00	24	7.88	3.49	0.00	14.00	184	8.90	4.34	1.00	16.00
	lagentry	208	0.14	0.35	0.00	1.00	24	0.21	0.41	0.00	1.00	184	0.13	0.34	0.00	1.00
	lagHHI	208	0.29	0.1	0.05	0.31	24	0.12	0.1	0.1	0.31	184	0.31	0	0.07	0.31
1998	lagsales	208	5.97	1.78	2.53	10.23	12	6.04	1.65	2.99	8.32	196	5.97	1.79	2.53	10.23
	lagsub	208	8.82	4.25	1.00	16.00	12	8.33	3.77	2.00	15.00	196	8.85	4.29	1.00	16.00
	lagentry	208	0.12	0.32	0.00	1.00	12	0.33	0.49	0.00	1.00	196	0.10	0.30	0.00	1.00
	lagHHI	208	0.28	0.1	0.06	0.29	12	0.11	0.1	0.1	0.29	196	0.29	0	0.07	0.29

Source: My elaboration on IMS Health data

4. The Econometric model

The aim of this paper is to emphasize the role played by submarket concentration in shaping a company's decision to enter at least one new submarket.

The model of entry probability in at least one new submarket is conditioned by whether entry occurred or not in the previous period.

A dichotomous probit model¹² is combined with the panel data structure of the dataset, in which the dichotomous dependent variable is constructed as $y_{it} = 1$ when company i at time t achieves zero sales in at least one of the submarkets in which its sales were zero at time $t - 1$; otherwise $y_{it} = 0$.

The conditional probability of entry is modeled on the basis of unit-specific regressors: the company's size in the previous period, the number of submarkets in which it previously operated; the earlier entry decision, (the lagged dependent variable) and the lagged submarket concentration level faced by companies assessing entry decisions. In addition, unit heterogeneity is accounted for with unit-specific intercepts (random effects).

This analysis adopts a Bayesian approach. The use of Bayesian inferential techniques is fairly rare in applied industrial economics¹³ but enables the inclusion of not strictly exogenous variables among the regressors, which can be correlated with unit heterogeneity. Several authors address this problem in a non-linear panel data framework, in particular: Amisano and Giorgetti (2013a); Arellano and Carrasco (2003); Honoré and Kyriazidou (2000); Honoré and Lewbel (2002); Wooldridge (2005).

More specifically, the probit model¹⁴ can be written as follows:

$$P(y_{it} = 1 | I_{t-1}, \theta) = \Phi(\alpha_i + \beta' z_{it-1}) \quad [1]$$

$$\beta = [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \gamma_1, \gamma_2, \gamma_3] \quad [2]$$

$$z_{it-1} = \left[x'_{it-1}, x^{(2)'}_{i0} \right] \quad [3]$$

$$x_{it-1} = [SA_{it-1}, SU_{it-1}, y_{it-1}, HHI_{it-1}, 1] \quad [4]$$

$$x_{i0}^{(2)'} = [SA_{i0}, SU_{i0}, y_{i0}] \quad [5]$$

where the dependent variable, y_{it} is a zero/one dummy that equates to the value of one when there is an entry into a new submarket at time t , while I_{t-1} is the information set that can include the past values of strictly exogenous regressors and the past and initial value of predetermined variables, θ is the vector of free parameters in the model.

¹² A clear explanation of sensible alternative choices can be found in Train (2003). Of course, it would also be possible to adopt a multinomial discrete choice approach, but in this particular case could seriously complicate both the estimation and the interpretation of the results.

¹³ The Bayesian approach is used more frequently in conjunction with financial or macroeconomic analysis.

¹⁴ Not strictly exogenous covariates as well as predetermined regressors can be incorporated.

The vector x_{it-1} includes the covariates suggested by economic considerations: SA_{it-1} , the sales of each company at time $t - 1$; SU_{it-1} , the number of submarkets in which company i was active in the previous period (i.e., the number of submarkets with sales higher than zero); y_{it-1} , the lagged dependent variable, and H_{it-1} , the lagged HHI calculated at submarket level. The last element x_{it-1} is the intercept term.

The vector $x_{i0}^{(2)}$ includes the initial condition for predetermined variables: SA_{i0} , SU_{i0} , y_{i0} , which refer, respectively, to each company's sales, number of submarkets entered, and entry decision calculated at the start of the review period (1988).

As in earlier papers¹⁵, it is assumed that random effects are dependent on covariates and on the initial condition as proposed by Wooldridge (2005), unobservable heterogeneity is modeled by specifying a distribution for unit-specific intercepts conditional on the initial values and on the values of the covariates:

$$p(c_i | y_{i0}; \mathbf{X}_i, \theta) \quad [6]$$

Further details on this point can be found in Amisano and Giorgetti (2013a).

5. Results

The estimated model has $k = 8$ covariates (the five regressors and the initial values of lagged sales, the number of submarkets and entry). The parameters of the model to be estimated are the slope coefficients in the probit specification (the vector β) and the inverse variance of the conditional distribution of the intercept terms (h_α). All the results are summarized in tables 4a to 4i, which show the posterior mean, the posterior standard deviation and the bounds (dubbed as lower and upper) of the marginal 95% highest posterior density (HPD) confidence intervals. A 95% HPD interval for a parameter is defined as the 95% probability mass interval associated with the highest values of the posterior distribution of that parameter. The symbol (*) indicates the parameters with HPD intervals that do not show a zero value, used, rather crudely, to state the significance of the associated parameters. Table 4a contains the aggregate data and tables 4b-4i the dataset evidence for the specific countries. Tables 4a to 4i also include measures of fit of the estimated models.¹⁶ All cases show high measures of fit, indicating the model's ability to explain the sample behavior: as an example, for the aggregate dataset, the two computed goodness-of-fit statistics are respectively point eighty-five and point ninety-six (see bottom of table 4a).

In order to provide a measure of the importance of single covariates in determining entry probabilities, all the results tables also show the average partial effects (APEs) and elasticities.¹⁷

In terms of the role of the regressors, the parameter β_1 for size, measured by lagged sales (*lagsales*), is not significant on an aggregate level (table 4a) nor it is significant for all the countries, with the main exceptions being France (see table 4c), Spain (see table 4f), and the

¹⁵ See: Amisano and Giorgetti (2008, 2013a, 2013b).

¹⁶ The two measures produced are: (i) the correlation between the dependent variable and the fitted probabilities. The fitted probabilities used are their posterior mean values. (ii) The sample proportion of correct "predictions": a correct "prediction" is when a low computed fitted probability, i.e., lower than the unconditional probability, corresponds to a zero value for the dependent variable, and vice versa. Once again, the posterior mean of the fitted probabilities are used.

¹⁷ For information on the way APEs are computed, see Amisano and Giorgetti (2013a). Elasticities are computed using the sample mean values of the covariates and of the entry probabilities.

U.S. (see table 4i). Company diversification is measured by the number of submarkets in which the company was present at time $t - 1$ β_2 , the coefficient of the lagged number of submarkets (*lagsub*), is not significant at an aggregate level (see table 4a) for most countries (see tables 4a to 4i) with Italy, Spain, and the U.S. being the exceptions, where *lagsub* has a negative and significant coefficient (see tables 4e, 4f and 4i).¹⁸ The effect of the achieved diversification is therefore sizeable for these countries. This negative relationship between achieved diversification and entry probabilities could be partly ascribed to a kind of saturation effect. Given that the number of submarkets is fixed (see section three), the number of unexploited entry possibilities for each company (our dependent variable) is negatively correlated to the number of submarkets already entered. However this result could be also explained by the presence of an “optimal core” of submarkets to enter.

The coefficient β_3 of the lagged dependent variable (*lagentry*) is not significant at the aggregate level (see table 4a). The finding of an insignificant role for the lagged dependent variable is confirmed at the disaggregated level, with the sole exceptions being France and Italy (see tables 4c and 4e). In the French case, the posterior distribution of β_3 assigns relevant probability mass to negative values, while the parameter β_3 is positive for Italy. This different sign for the parameter β_3 is quite hard to explain but could be caused by the significance of two regressors in France that are not significant in Italy: the positive sign of the size at the initial period of observation and the lagged company size with a negative sign.

In addition, the different results could be due to the significance in France of the level of submarket diversification already reached by the sample in the initial review period, the significance (with a negative sign) of the lagged sales in France and the insignificant role of the lagged number of submarkets. Thus, future research will aim to measure the distribution of the companies’ global sales among the submarkets because it could potentially explain the conflicting results between France and Italy on the sign of the lagged dependent variable.

In terms of the role of submarket concentration on entry decisions, the concentration level, as measured by the *lagHHI* regressor, is significant and negative. At an aggregate level (table 4a), the posterior mean of the coefficient associated with *lagHHI*, β_4 is significant, negative and very high (- 14.26), corresponding to an APE of - 1.15 and an elasticity of - 2.76. This very strong result is confirmed by the outcome produced by the single countries dataset. With no exceptions, the β_4 parameter is always significant and negative (see tables 4b to 4j). It can also be noted that the elasticities corresponding to the posterior means of β_4 are quite stable (between - 2 and - 3) for most countries, with the exception of Canada, which has an estimated elasticity of around -6. These results are quite clear as they all indicate the relevant role of concentration as an entry barriers: high concentration does not drive diversification but clearly acts as a barrier. A potential explanation of the negative role of submarket concentration on entry decisions could be the recent problems experienced by the pharmaceutical companies in the export of long-term first-mover advantages to different classes of products due to the entry barriers of the largest incumbents faced by the potential entries in each submarket. Barbosa (2003), without differentiating for entry types, finds a negative relationship between entry and market concentration. Thus, the results for entry by diversification at the submarket level

¹⁸ The posterior estimated means for the β_2 coefficient are, respectively: -.24, -.41 and -.19.

(without differentiating for entry types) replicate the conclusions of the literature on the relationship between entry and market concentration.

Table 4.a. Probit estimation results. all countries. n=208

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales	0,032	0,083	-0,132	0,033	0,193	0,002	0,109
β_2	lagged submarkets	-0,043	0,045	-0,133	-0,043	0,043	-0,003	-0,231
β_3	lagged entry	0,174	0,168	-0,161	0,175	0,498	0,013	0,013
β_4	lagged HHI*	-14,263	0,625	-15,515	-14,252	-13,069	-1,114	-2,763
β_5	Intercept*	2,512	0,257	2,018	2,508	3,024	-	-
β_6	initial sales	-0,078	0,086	-0,246	-0,078	0,092	-0,006	-0,252
β_7	initial submarkets	0,063	0,050	-0,032	0,062	0,162	0,005	0,321
β_8	initial entry	0,192	0,417	-0,636	0,194	0,998	0,015	0,004
h_9	- *	2,346	0,506	1,493	2,302	3,464	-	-
goodness of fit								
correlation between p^{\wedge} and y		0,854						
proportion correct forecasts		0,964						

Table 4.b. Probit estimation results. Canada. n=71

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales	-0,298	0,189	-0,681	-0,291	0,055	-0,051	-1,272
β_2	lagged submarkets	0,074	0,076	-0,073	0,074	0,224	0,013	0,547
β_3	lagged entry	-0,047	0,255	-0,561	-0,043	0,440	-0,008	-0,007
β_4	lagged HHI*	-12,477	1,367	-15,443	-12,373	-10,104	-2,151	-6,087
β_5	Intercept*	4,902	0,755	3,560	4,854	6,532	-	-
β_6	initial sales	0,090	0,196	-0,279	0,084	0,488	0,015	0,349
β_7	initial submarkets	0,017	0,083	-0,146	0,017	0,181	0,003	0,118
β_8	initial entry	-0,013	4,476	-8,748	-0,020	8,730	-0,003	0,000
h_9	- *	3,022	1,262	1,244	2,786	6,079	-	-
goodness of fit								
correlation between p^{\wedge} and y		0,868						
proportion correct forecasts		0,930						

Table 4.c. Probit estimation results. France. n=75

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales *	-0,774	0,199	-1,181	-0,768	-0,401	-0,091	-4,411
β_2	lagged submarkets	-0,008	0,069	-0,145	-0,008	0,126	-0,001	-0,057
β_3	lagged entry *	-0,492	0,252	-1,000	-0,486	-0,013	-0,058	-0,056
β_4	lagged HHI*	-7,791	0,721	-9,232	-7,779	-6,414	-0,915	-2,543
β_5	Intercept*	2,112	0,547	1,060	2,104	3,207	0,248	
β_6	initial sales*	0,613	0,185	0,267	0,606	0,991	0,072	3,298
β_7	initial submarkets	0,005	0,074	-0,139	0,004	0,151	0,001	0,036
β_8	initial entry	0,012	4,456	-8,718	0,016	8,799	0,001	0,000
h_n	- *	2,972	1,116	1,373	2,780	5,681	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,707						
proportion correct forecasts		0,789						

Table 4.d. Probit estimation results. Germany. n=73

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales	-0,131	0,137	-0,414	-0,127	0,124	-0,019	-0,885
β_2	lagged submarkets	-0,008	0,059	-0,126	-0,007	0,107	-0,001	-0,087
β_3	lagged entry	0,040	0,221	-0,403	0,044	0,463	0,006	0,005
β_4	lagged HHI*	-8,699	0,657	-9,993	-8,696	-7,414	-1,245	-2,855
β_5	Intercept*	1,337	0,338	0,671	1,338	1,994	0,192	
β_6	initial sales	0,007	0,140	-0,255	0,002	0,292	0,001	0,036
β_7	initial submarkets	0,053	0,068	-0,077	0,052	0,189	0,008	0,525
β_8	initial entry	0,003	1,003	-1,968	0,004	1,961	0,001	0,000
h_n	- *	5,050	1,725	2,409	4,803	9,071	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,720						
proportion correct forecasts		0,788						

Table 4.e. Probit estimation results. Italy. n=71

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales	0,095	0,156	-0,213	0,096	0,401	0,013	0,475
β_2	lagged submarkets*	-0,241	0,101	-0,451	-0,237	-0,053	-0,034	-1,833
β_3	lagged entry *	0,432	0,212	0,014	0,433	0,844	0,061	0,058
β_4	lagged HHI*	-9,995	0,748	-11,507	-9,979	-8,571	-1,425	-2,917
β_5	Intercept*	1,271	0,594	0,116	1,266	2,450	0,182	
β_6	initial sales	0,058	0,167	-0,267	0,057	0,392	0,008	0,285
β_7	initial submarkets	0,189	0,098	0,006	0,185	0,391	0,027	1,406
β_8	initial entry	-0,015	4,469	-8,765	-0,024	8,742	-0,002	0,000
h_n	- *	3,228	1,318	1,327	3,000	6,391	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,774						
proportion correct forecasts		0,777						

Table 4.f. Probit estimation results. Spain.n=81

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	Lagged sales *	1,052	0,401	0,283	1,046	1,853	0,040	1,800
β_2	lagged submarkets*	-0,408	0,167	-0,735	-0,408	-0,083	-0,015	-1,129
β_3	lagged entry	0,219	0,512	-0,850	0,244	1,154	0,008	0,008
β_4	lagged HHI*	-13,159	1,752	-16,791	-13,089	-9,986	-0,496	-3,234
β_5	Intercept*	4,461	1,232	2,188	4,412	7,009	0,171	
β_6	initial sales	-0,600	0,392	-1,361	-0,604	0,183	-0,023	-0,980
β_7	initial submarkets	0,311	0,175	-0,034	0,312	0,654	0,012	0,859
β_8	initial entry	0,006	4,469	-8,805	0,039	8,728	0,000	0,000
h_a	- *	2,170	1,143	0,686	1,925	5,050	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,938						
proportion correct forecasts		0,982						

Table 4.g. Probit estimation results. UK. n=70

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales	0,103	0,133	-0,158	0,104	0,366	0,012	0,444
β_2	lagged submarkets	-0,146	0,091	-0,327	-0,144	0,030	-0,017	-0,872
β_3	lagged entry	-0,674	0,381	-1,472	-0,655	0,022	-0,079	-0,071
β_4	lagged HHI*	-13,776	1,301	-16,494	-13,718	-11,391	-1,628	-5,302
β_5	Intercept*	4,521	0,632	3,349	4,497	5,823	0,535	
β_6	initial sales	-0,160	0,122	-0,399	-0,161	0,082	-0,019	-0,621
β_7	initial submarkets	0,168	0,092	-0,008	0,167	0,354	0,020	0,988
β_8	initial entry	-0,025	4,462	-8,777	-0,029	8,715	-0,003	0,000
h_a	- *	3,141	1,288	1,326	2,908	6,280	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,869						
proportion correct forecasts		0,925						

Table 4.h. Probit estimation results. USA. n= 57

coef	regressors	mean	s.d.	lower	median	upper	APE	elasticity
β_1	lagged sales *	0,201	0,046	0,112	0,200	0,292	0,053	1,013
β_2	lagged submarkets*	-0,193	0,049	-0,293	-0,193	-0,099	-0,051	-1,327
β_3	lagged entry	0,254	0,162	-0,062	0,254	0,572	0,067	0,062
β_4	lagged HHI*	-5,129	0,410	-5,957	-5,123	-4,346	-1,357	-2,669
β_5	Intercept*	1,423	0,261	0,915	1,421	1,938	0,377	
β_6	initial sales	-0,096	0,061	-0,217	-0,096	0,022	-0,026	-0,335
β_7	initial submarkets *	0,169	0,046	0,081	0,168	0,263	0,045	1,099
β_8	initial entry	0,009	4,476	-8,740	0,003	8,753	0,002	0,000
h_a	-	4,083	1,433	1,958	3,851	7,512	0,000	
goodness of fit								
correlation between p^{\wedge} and y		0,73						
proportion correct forecasts		0,63						

Table 4.i. Probit estimation results. Reduced models, coefficient on HHI

coef	mean	s.d.	lower	median	upper	APE	elasticity
All	-10,783	0,415	-11,608	-10,777	-9,983	-1,115	-2,764
Canada	-11,985	1,297	-14,807	-11,889	-9,708	-2,215	-5,494
France	-7,754	0,709	-9,189	-7,738	-6,403	-0,917	-2,273
Germany	-19,178	1,580	-22,434	-19,121	-16,244	-2,336	-5,793
Italy	-9,907	0,743	-11,413	-9,893	-8,497	-1,432	-3,551
Spain	-13,475	1,786	-17,255	-13,369	-10,350	-0,446	-1,107
UK	-13,218	1,260	-15,883	-13,147	-10,968	-1,714	-4,250
USA	-5,064	0,405	-5,870	-5,057	-4,292	-1,341	-3,325

Source: my elaboration on IMS Health data

Conclusion

The paper has investigated the impact of concentration on entry into a new submarket, by which we mean the role of submarket concentration when entry happens through diversification. The impact of “market concentration” changes according to the way companies decide to enter (Mueller, 1991). Precisely the process of entry by diversification at market level is producing conflicting results, and the role of concentration from the submarket perspective has not been well analyzed. The paper uses a dynamic Bayesian panel probit to investigate whether the probability of entry into a new submarket at time t , conditioned by entry or non-entry into a new submarket in the previous period, is positively or negatively connected to the role of submarket concentration.

The paper investigates entry into new submarkets in an aggregate of seven countries (Canada, France, Italy, Germany, Spain, the UK, and the U.S.) and seeks to capture each country’s idiosyncrasies (as suggested in the pharmaceutical sector literature, Kyle, 2006) by repeating the same analysis for each country separately. The analysis takes into account company-specific characteristics, i.e., lagged sales, lagged entry, the number of lagged submarket entries, and market specific regressors (namely submarket specific regressors) as the lagged concentration in each submarket.

The analysis asks an important research question: what impact does submarket concentration have on the decision to enter a new submarket? The findings suggest there is no incentive for companies to enter submarkets dominated by a high level of (sub-) market power and that entry therefore correlates negatively to concentration at the submarket level.

However, the study has generated other important results for the seven countries analyzed, both severally and jointly, highlighting important country-specific peculiarities: in Italy, Spain, and the U.S., the number of submarkets already entered has a negative impact on entry; in Spain and the U.S., lagged size is positive but negative in Italy, preventing confirmation of whether a positive relationship exists between size and future diversification decisions. Lagged entry is significant and positive in Italy but significant and negative in France. Overall, the principal aim of this paper is to investigate the relationship between entry by diversification and concentration by adopting a more submarkets-based framework (Sutton, 1998). The future research pathway will analyze entry decisions using a three digit submarket classification to shed light on the role of concentration in each specific submarket, adopting a standard entry-exit framework with demand factors, cost factors, and degree of competition separately for each specific submarket.

Akerberg D., Benkard L., Berry S. and Pakes A. (2007). Econometric tools for analyzing market outcomes. *Handbook of Econometrics*, 6: 4171-4276.

Aguiregabirria Vand Mira P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75 (1): 1-53.

Amisano G. and Giorgetti M.L. (2008). *Entry into New Pharmaceutical Submarkets: The Role of Submarket Concentration Electronic*, published online.

Amisano G. and Giorgetti M.L. (2013a). Diversification by entry into a new submarket? *Applied Economics*, vol 45, issue 12, pp.1507-1518.

Amisano G. and Giorgetti M.L. (2013b). Entry in Pharmaceutical Submarkets: A Bayesian Panel Probit Analysis, *Journal of Applied Econometrics*, Vol. 28, No. 4, 2013, pp. 667-701.

Arellano M. and Carrasco R. (2003). Binary choice panel data models with predetermined variables. *Journal of Econometrics*, 115 (1): 125-157.

Baldwin J.R. (1995). *The Dynamics of Industrial Competition: A North American Perspective*. Cambridge University Press: Cambridge (UK).

Barbosa N. (2003). What drives new firms into an industry? An integrative model of entry. *NIMA (Núcleo de investigação em micro-economia aplicada) Working Papers 23*

Berry C.H. (1975). *Corporate Growth and Diversification*. Princeton University Press: Princeton.

Berry S. and Reiss P. (2007). Empirical models of entry and market structure, in Armstrong M., Porter R. (eds.) *Handbook of Industrial Organization*. Elsevier: Amsterdam.

Biewen M. (2004). Measuring state dependence in individual poverty status: are there feedback effects to employment decisions and household compositions. *IZA Discussion Paper 1138*.

Bottazzi G., Dosi G., Lippi M., Pammolli F. and Riccaboni M. (2001). Innovation and corporate growth in the evolution of the drug industry. *International Journal of Industrial Organization*, 19 (7): 1161-1187.

Caves R.E, Whinston M.D. and Hurwitz M.A. (1991). *Patent Expiration, Entry, and Competition in the U. S. Pharmaceutical Industry*. The Brookings Institution: Washington.

Danzon P.M., Wang Y.R. and Wang L. (2005). The impact of price regulation on the launch delay of new drugs. Evidence from twenty-five major markets in the 1990s. *Health Economics*, 14: 269-292.

Deutsch L. (1975). Structure, performance and the net rate of entry in manufacturing industries. *Southern Economic Journal*, 41: 450- 456.

Di Masi J. (2001). Winners and losers in new drug innovation. *Medical Marketing & Media*, 36 (9): 99-110.

Dosi G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11 (3): 147-162.

Ellickson P., Houghton S. and Timmins C. (2007). Estimating network economies in retail chains: a revealed preference approach. *NBER Working Paper w15832*.

Geroski P.A. (1999). The growth of firms in theory and practice. *CEPR Discussion Paper 2092*

Geweke J. and Keane M.P. (1999). Mixture of normals probit models, in Hsiao C., Lahiri K., Lee L.F., Pesaran H. (eds.) *Analysis of Panels and Limited Dependent Variable*

Models: An Edited Volume in Honor of G.S. Maddala. Cambridge University Press: Cambridge, UK.

Giorgetti M.L. (2006). Scope economies in the pharmaceutical sector. *International Review of Economics*, 53 (3): 373-401.

Henderson R. and Cockburn I. (1996). Scale, scope, and spillovers: the determinants of research productivity in drug discovery. *Rand Journal of Economics*, 27 (1): 32-59.

Holmes T. (2011). Diffusion of Wal-Mart and economies of density. *Econometrica*, 79 (1): 253-302.

Honoré B.E. and Kyriazidou E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68 (4): 839-874.

Honoré B. and Lewbel A. (2002). Semiparametric binary choice panel data models without strictly exogenous regressors. *Econometrica*, 70 (5): 2053-2063.

Klepper S. and Thompson P. (2006). Submarkets and the evolution of market structure. *Rand Journal of Economics*, 37 (4): 861-886.

Kyle M. (2006). The role of firms characteristics in pharmaceutical product launches. *Rand Journal of Economics*, 37 (3): 602-618.

Lanjouw J. (2005). Patents, price controls, and access to new drugs: how policy affects global market entry. *NBER Working Paper 11321*.

MacDonald J.M. (1984). Diversification, market growth and concentration in U.S. manufacturing. *Southern Economic Journal*, 50 (4): 1098-1111.

Matraves C. (1999). Market structure, R&D and advertising in the pharmaceutical industry. *Journal of Industrial Economics*, 47 (2): 169-194.

Mueller D. (1991). Entry, exit and the competitive process, in Geroski P., Schwalbach J. (eds.) *Entry and Market Contestability*. Basil Blackwell: Oxford.

Panle J. (2008). What happens when Wal-Mart comes to town: an empirical analysis of the discount retailing industry. *Econometrica*, 76 (6):1263-1316.

Penrose E. (1959). *The Theory of Growth of the Firm*. Basil Blackwell: Oxford.

Reiss P. and Wolak F. (2007). Structural econometric modeling: rationales and examples from industrial organization, in Armstrong M., Porter R. (eds.) *Handbook of Industrial Organization*. Elsevier: Amsterdam.

Scott Morton F. (1999). Entry decision in the generic pharmaceutical industry. *Rand Journal of Economics*, 30 (3): 421-440.

Sutton J. (1998). *Technology and Market Structure*. MIT Press: Cambridge (Mass.).

Train K.E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press: Cambridge (UK).

Wooldridge J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, 20 (1): 39-54.

Table A1 – First level classification including sixteen submarkets

1. A =alimentary tract and metabolism products
2. B =blood and blood forming organs
3. C =cardiovascular system products
4. D =dermatological products
5. G =genito-urinary system and sex hormone products
6. H =systemic hormonal preparations (excluding sex hormone) products
7. J =general anti-infective-systemic products

8. K =hospital solutions
9. L =antineoplastic and Immunomodulating agent products
10. M =musculo-skeletal system products
11. N =central nervous system products
12. P =parasitology products
13. R =respiratory system products
14. S =sensory organs products
15. T =diagnostic agents
16. V =various

Source: my elaboration on IMS Health data