UNIVERSITA' DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA GESTIONALE E DELL'INFORMAZIONE[°]
QUADERNI DEL DIPARTIMENTO[†]

Department of Management and Information Technology

Working Paper

Series "*Mathematics and Statistics*"

n. 6/MS – 2006

Multivariate statistical sensitivity analysis of a computer model for pharmaceutical industry market and innovation dynamics

by

Alessandro Fassò, Franco Malerba, Luigi Orsenigo, Michele Pezzoni

[°] Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779
[†] Il Dipartimento ottempera agli obblighi previsti dall'art. 1 del D.L.L. 31.8.1945, n. 660 e successive modificazioni.

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* ed i *Technical Reports* della Collana dei Quaderni del Dipartimento di Ingegneria Gestionale e dell'Informazione costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Multivariate statistical sensitivity analysis of a computer model for pharmaceutical industry market and innovation dynamics

Alessandro Fassò*, Franco Malerba†, Luigi Orsenigo‡, Michele Pezzoni§

September 22, 2006

Abstract

Statistical sensitivity analysis is a useful technique to analyze multivariate stochastic computer model in order to better understand the cause and effect relationships between input parameters and output observations. This paper is based on a pre-existent formal evolutionary economic model that simulates the main aspects of the market and the innovation processes that take place inside the pharmaceutical industry. It belongs to the family of 'History-Friendly' models. Our purpose is to reveal the critical input parameters concerning R&D costs, research opportunities, regulatory regime, demand and firm's features in the mechanisms of innovation and market dynamics through the use of multivariate statistical sensitivity analysis. This preliminary work represents a first step in the introduction of a complete analysis with a mixed linear model.

Keywords: History-Friendly model, Pharmaceutical industry, Stochastic computer model, Mixed linear model.

*University of Bergamo, Dept. IGI, Via Marconi 5, 24044 Dalmine BG, Italy.
Email: alessandro.fasso@unibg.it. Work partially supported by Italian MIUR-Cofin 2004 grants.

†CESPRI, Bocconi University, Milan, Italy.
Email: franco.malerba@unibocconi.it

‡Department of Engineering, University of Brescia and CESPRI, Bocconi University, Milan, Italy.
Email: luigi.orsenigo@unibocconi.it

§Department of Engineering, University of Brescia, Italy.
Email: michele.pezzoni@tin.it

1 Introduction

In this paper we propose a concise method to analyze a stochastic computer model of the pharmaceutical industry through the use of the sensitivity analysis technique. The model belongs to the family of 'History-Friendly' evolutionary economic models. These models aim to capture the essence of the qualitative theory, pointed out by the technology and industry scholars. They also try to give a possible logical explanation. The modelling process philosophy starts from empirical studies which describe the mechanisms inside pharmaceutical industry. The motivations underlying this modelling stile has been discussed extensively in previous works, see e.g. Malerba et al.,(1999),(2001). Is worth to remember that one of the reason for interest in History-Friendly models is to understand what kind of factors and dynamic processes are important for the industry evolution. In this paper we try to explain what are the factors that can indeed explain the observed patterns of industrial dynamics, in particular about innovation and imitation processes, prices and industry structure. The technique that appears more promising for assessing the impact of the main industry parameters, seems to be the sensitivity analysis (SA). The modern concept of sensitivity analysis for computer models arises from the logical intersection of some variance decomposition and some sampling or computer experiment techniques, see e.g. Saltelli et al. (2000) and references therein. This technique has been recently applied to various kinds of computer models both in Econometrics and Finance, see e.g. Saltelli et al (2004) and in environmental models, see e.g. Fassò (2006) and Fassò and Perri (2002) . Cheap computer models, which run fast, and expensive computer models, which take CPU time and storage capability, usually requires different SA techniques. In this paper, we are concerned with a cheap code based extensive simulations which allow to assess the role and the interactions of the various input parameters in influencing the multi-output model. The paper is organized as follows. Section 2 introduces the probabilistic structure of a Stochastic Computer Model (SCM) for assessing the uncertainty arising from the so-called stochastic game and proposes a statistical emulator to be estimated on available simulated data. Section 3, after extending sensitivity analysis (SA) to cope with SCM, discusses appropriate sampling and preliminary estimation techniques for the proposed SCM emulator. Section 4, after a brief summary of the pharmaceutical industry history, introduces the key features of the Stochastic Computer Model. In Section 5 we assess the influence of input parameters which concern costs, environment and opportunity conditions, regulatory regime, firm's features and demand on multi-outputs in innovation and imitation processes, prices and industry structure.

2 Stochastic Computer Models

A Stochastic Computer Model (SCM) may be described by a computable code or multi-valued function z such that

$$z = z(x, \xi),$$

where x is a k-dimensional input parameter which is known at the time the code is run, while ξ is an unobservable random vector, which is generated by the code itself. In market behavior SCM's, the random element ξ is called the stochastic game component of the SCM and the input model $x = (x_1, ..., x_k)$ defines e.g. market dimensions, side conditions, initial values.

In the setup of SA, x has a known probability distribution, $P(x)$ say, which may be a multivariate distribution with discrete and/or continuos marginals and possible dependency structure among the x components. Generally speaking, the stochastic game ξ has a known or unknown conditional distribution

$$P(\xi|x)$$

and, in some cases, it is independent on x, so that $P(\xi|x) = P(\xi)$.

It turns out that, for given x, the output z is a random vector with distribution

$$P(z|x)$$

with the first two conditional moments given by

$$\mu(x) = E(z|x)$$

and

$$\tau^2(x) = Var(z|x).$$

The output uncertainty is given, in general terms, by the unknown marginal distribution

$$P(z)$$

and this may be decomposed into two components, one is related to the stochastic game ξ and the other one accounts for the input variability.

When the output z is single-valued, using $E()$ for the mathematical expectation and $Var()$ for the variance of a random variable, recalling the well known ANOVA decomposition, which is given by

$$Var(z) = Var(\mu(x)) + E\left(\tau^2(x)\right), \tag{1}$$

we note that the rate

$$\frac{E(\tau^2(x))}{Var(z)}$$

is the residual uncertainty or the quota of the global output uncertainty which depends on the stochastic game ξ averaged over the input space.

On the other side, $Var(\mu(x))$ is the accountable output uncertainty and may be related to each single input x_j using SA techniques discussed in section 3 below. Actually

$$\eta^2 = \frac{Var(\mu(x))}{Var(z)}$$

is the well known Pearson's correlation ratio and $1 - \eta^2$ represents the best fitting error among emulators which will be discussed below.

2.1 Heteroskedastic SCM's

We say that the SCM is conditionally homoskedastic if the conditional output uncertainty does not depend on the particular input parameter value $x = x_0$, namely if

$$\tau^2(x) = \tau^2$$

for every x. Otherwise the SCM is called conditionally heteroskedastic and $\tau^2(x)$ gives the conditional or local output uncertainty whose understanding and modelling gives further insight into SCM comprehension. For example, Fassò et al. (2003) considered a parabola model $\tau^2(x) = \alpha_0 + \alpha_1 x_1^2$ based only on the first input x_1.

2.2 Input feasibility set

Let D_z be the (extended) output set and let $F_z \subset D_z$ be the output feasibility set where the model give pleasant outputs. For example we exclude from F_z those points corresponding to a zero firms market (all died).

It is then of interest to define the input feasibility set F_x, which is such that the corresponding average output is feasible, in symbols:

$$x \in F_x \Rightarrow \mu(x) \in F_z. \tag{2}$$

In this case it may be important to compute the probability of an unfeasible output, namely

$$u(x) = P(z \in D_z - F_z | x)$$

for every feasible input $x \in F_x$.

Note that feasibility concept may introduces constraints on the input parameter distribution which are not compatible with input independence. For example one may start from an unconstrained input distribution $P_0(x)$ with independent components,

$$P_0(x) = P_0(x_1) \cdot ... \cdot P_0(x_k)$$

but after imposing feasibility, that is using

$$P(x) = P_0(x | x \in F_x)$$

such independence may be lost.

2.3 SCM emulator

One way of attaching the understanding of z may be based on the so-called mixed linear models. According to this we suppose that the influence of the input parameters x and the the stochastic game ξ may be separated and the influence of x may be modelled by an appropriate *emulator* which is a simplified statistical model. In particular the output z, possibly after an appropriate transformation, *e.g.* log-transform, is defined as follows:

$$z = z(x, \xi) = f(x) + \varepsilon(\xi)\,\sigma(x) \tag{3}$$

where

$$\begin{aligned} f(x) &= g(x, \beta) + e \\ &= \beta_0 + \sum_{j=1}^{p} g_j(x)\,\beta_j + e \end{aligned} \tag{4}$$

and

$$\sigma^2(x) = \alpha_0 + \sum_{j=1}^{q} h_j(x)\,\alpha_j \tag{5}$$

Here the stochastic game standardized component $\varepsilon = \varepsilon(\xi)$ has a zero mean and unit variance Gaussian noise which does not depend on x and is also independent on the emulator error e. Similarly, the latter is supposed a Gaussian noise with zero mean and variance σ_e^2. Both function sets $g_j()$ and $h_j()$ may include various input components such as the linear-, quadratic-, interaction- etc.

From equation (4) , we see the role of the emulator $g()$ which is a simplified model for f with emulation error e and coefficient vector $\beta = (\beta_0, ..., \beta_p)$ to be estimated.

Similarly the skedastic function of equation (5) is identified by the coefficient vector $\alpha = (\alpha_0, ..., \alpha_q)$ where, in particular, α_0 is related to the stochastic game component which does not depend on x. and the other terms α_j are related to the interactions among the stochastic game and the input components $h_j(x)$.

For non Gaussian SCM's, for example when z is a counting variable, we can use a generalized mixed linear model, see e.g. Lee and Nelder (2001).

3 Sensitivity Analysis for SCM

The model of section 2.3 is useful for SCM understanding. From the practical point of view it is often useful to assess the various sources of uncertainty by ranking the input x_j according to their influence on the output uncertainty.

Input component		Sensitivity
linear	$g_1(x) = x_1$	$S_1 = \beta_1^2 Var(x_1)/Var(z)$
quadratic	$g_2(x) = x_1^2$	$S_2 = \beta_2^2 Var(x_1^2)/Var(z)$
interaction	$g_3(x) = x_1 x_2$	$S_3 = \beta_3^2 Var(x_1 x_2)/Var(z)$
Stochastic game	$h_1(x) = x_1^2$	$HS_1 = \alpha_1 E(x_1^2)/Var(z)$
Total sensitivity to x_1		$S_{T_1} = S_1 + S_2 + S_3 + HS_1$

Table 1: Example of sensitivity decomposition

To do this we can apply the approach of equation 1 to the above mentioned model. If the model components $g_1(x), ..., g_p(x), \varepsilon$ are uncorrelated, then, extending the heteroskedastic SA of Fassò et al. (2003) for SCM's, we get the following $ANOVA$ decomposition

$$Var(z) = \sum_{j=1}^{p} \beta_j^2 Var(g_j(x)) + \alpha_0 + \sum_{j=\backslash}^{q} E(h_j(x))\alpha_j + \sigma_e^2. \tag{6}$$

It follows a natural definition of sensitivity, namely the sensitivity of the input component $g_j(x)$ is given by

$$S_j = \beta_j^2 \frac{Var(g_j(x))}{Var(z)} \tag{7}$$

and the heteroskedastic component $h_j(x)$ gives the following sensitivity index:

$$HS_j = \alpha_j \frac{E(h_j(x))}{Var(z)}. \tag{8}$$

Natural ranking and grouping of the relevant input components may then be performed using the sensitivity indexes.

Example 1 *As a first example consider an homoskedastic SCM with a simple linear emulator and orthogonal inputs*

$$g(x, \beta) = \beta_0 + \sum_{j=1}^{k} x_j \beta_j$$

then $S_j = \beta^2 Var(x_j)/Var(z)$ can be sorted and displayed in tabular format giving a simple synthesis of the SA. Also note that, in this case, $\sum S_j = 1 - R^2$, where R is the well-known multiple linear correlation coefficient.

Example 2 *As a second example, consider a model such that the first input x_1 enters equation (4) with linear, quadratic and one interaction component as in Table 1, then S_{T_1} is the total effect of x_1 on z and $S_{T_1} - HS_1$ is the total effect of x_1 keeping the stochastic game as fixed.Note that for homoskedastic models*

$$\sum_{j=1}^{p} S_j = 1 - \frac{\alpha_0 + \sigma_e^2}{Var(z)} \tag{9}$$

and in the case of the standard regression result reported above.

3.1 Multivariate SA

In many cases the output of a SCM is multivariate, namely $z = (z_1, ..., z_r)$, where the correlation among the outputs z_j is described by the variance-covariance matrix $V_z = Var(z)$.

To cope with this case, we consider the sensitivity index of Fassò (2006*b*), which is based on the variance decomposition of the linear combination $\alpha' z$. To see this, suppose that the *emulator* is in the form

$$g(x, \beta) = Bx = \sum_{j=1}^{k} B_j x_j$$

where B is a $r \times k$ matrix with j^{th} column B_j and

$$\alpha' V_z \alpha = \alpha' B V_x B' \alpha + \alpha' V_\varepsilon \alpha.$$

Hence, if x has uncorrelated components with $V_x = diag\left(\sigma^2_{x_1}, ...\sigma^2_{x_k}\right)$, the *quadratic* sensitivity index

$$S_j^\alpha = \sigma^2_{x_j} \frac{\alpha' B_j B_j' \alpha}{\alpha' V_y \alpha} \tag{10}$$

takes into account the correlation among the components of z and retains additivity as $R^2_{\alpha' z} = 1 - \frac{\alpha' V_\varepsilon \alpha}{\alpha' V_z \alpha} = \Sigma_j S_j^\alpha$. When inputs are correlated the above decompositions may still hold approximately. Otherwise orthogonalization techniques are available.

3.2 Monte Carlo approach

In this section we consider the peculiar aspect of SA for SCM, namely the data production process. The idea is to get m input samples $x^i = (x_1^i, ..., x_k^i)$, then run the computer code getting a model output y^i. These input-output data (x^i, z^i), $i = 1, ..., m$, can be used to get estimates of model (3) and sensitivity indexes of equation (7) and (8).

In order to get input samples we may follow two main ways, namely random and nonrandom sampling. The latter is based on design of (computer) experiment approach and is not considered anymore here.

The former approach starts from standard Monte Carlo which is simply based on m independent samples from the input distribution $P(x_1, ..., x_k)$. Various modifications are possible from improving model estimates based for example on Latin hypercube sampling, see e.g. McKey et al. (1979).

Note that, if we are interested in model based SA, an input distribution ensuring input orthogonality allows interpretation of indexes (7) and (8) as summable quantities for the particular model used. On the other hand this may be in contrast with the particular input feasibility set F_x used and some caution is in order. In some cases additivity still holds after suitably grouping the inputs.

In order to allow the estimation of the stochastic game, we need a stratified Monte Carlo sampling strategy. In our case this is simply done by n repeated computer runs for each fixed x^i. According to this, the input-output data are now given by

$$\left(x^i, z^{ij}\right),\ i = 1, ..., m,\ j = 1, ..., n$$

3.3 Preliminary analysis and model estimation

Whenever the practitioner may be interested primarily in input ranking irrespective of the stochastic game, indeed this component may be quite relevant both from the interpretative point of view, see e.g. equation (9), and from the model estimation point of view.

Thanks to our stratified approach, we can easily have a preliminary estimate of the minimum total model error

$$\tau^2(x) \cong \sigma^2(x) + \sigma_e^2$$

by

$$\hat{\tau}_i^2 = \hat{\tau}^2\left(x^i\right) = \frac{1}{n-1}\sum_{j=1}^{n}\left(y^{ij} - \hat{\mu}_i\right)^2$$

where $\hat{\mu}(x^i)$ is simply the average of the n replicates, namely

$$\hat{\mu}_i = \hat{\mu}\left(x^i\right) = \frac{1}{n}\sum_{j=1}^{n} y^{ij}$$

which is important for assessing heteroskedasticity.

3.3.1 Preliminary estimation

At the first stage, in order to estimate the input effect only, we can consider averaged outputs $\hat{\mu}_i$, $i = 1, ..., m$, where the stochastic game effect is reduced to $\frac{1}{n}\sigma^2(x)$.

In case where the averaged stochastic game component $\frac{1}{n}\sigma^2(x)$ is small with respect to the emulator error σ_e^2 we simply apply ordinary least squares to data $(\hat{\mu}_i, x^i)$. Otherwise, we use weighted least squares with covariance matrix given by

$$diag\left(\frac{\hat{\tau}_1^2}{n}, ..., \frac{\hat{\tau}_m^2}{n}\right).$$

In this way, we get preliminary estimates of β and perform SA as in section 3.

3.3.2 Mixed linear model

We can refine above estimates and get a full emulator identification by applying model (3) to the stratified Monte Carlo sample. This gives

$$z^{ij} = g\left(x^i, \beta\right) + \varepsilon^{ij}\sigma\left(x^i\right) + e^i$$

which is a linear model with mixed effects, where the input components $g_j(x^i)$ are the outer covariates, the emulator error e^i is the random effect and $\varepsilon^{ij}\sigma(x^i)$ is the heteroskedastic error.

4 A model for pharmaceutical industry

4.1 Main features of the market

In this section we give an overall description of pharmaceutical industry evolution resuming only in general therms the main patterns of development analyzed by several scholars.

The history of pharmaceutical industry can be usefully divided into three major epochs. The first, corresponding roughly to the period 1850-1945, in which little new drug development occurred, and in which there was very little research and was based on relatively primitive methods. The large scale of development of penicillin during the World War II marked the emergence of the second period of industry evolution. This period was characterized by the institution and formalization of in-house R&D programs and relatively rapid rates of drug introduction. During the early part of this period the industry relied largely on so called 'random' screening as method for finding new drugs, but in the seventies the industry began a transition to 'guided' drug discovery, a research methodology that allowed a great advance in molecular biochemistry, pharmacology and enzymology. The third epoch of the industry has its roots in the seventies but did not come to full flower until quite recently as the use of the tools of genetic engineering in the production and discovery of new drugs become more widely diffused.

An in-deep story and extensive analysis and discussion of the patterns of pharmaceutical industry have been undertaken by several scholars and will not be discussed here. In this paper we model the 'random' screening period but we report some hints to the whole story of the industry. As a way of introduction we briefly discuss the key aspects strictly linked to the model. In the history of the pharmaceutical industry, faced with such a 'target-rich' environment but with very little detailed knowledge of the biological underspinnings of specific diseases, pharmaceutical companies developed an approach to research that is now referred as 'random screening'. Under this approach, natural and chemically derived compounds are randomly screened in test-tube experiments and laboratory animals for potential therapeutic activity. Pharmaceutical companies maintained enormous 'libraries' of chemical compounds, and increased their collections by searching for new compounds in places such as swamps, stream and soil samples. Thousand of compounds might be subjected to multiple screens before researchers can focus on promising substance. Serendipity played a key role since in general the 'mechanism of action' of most drug was not well understood. Researchers were generally forced to rely on the use animal models as screens. Under this regime it was not uncommon for companies to discover a drug to treat one disease while searching for a treatment for another. The 'design' of new compounds

was a slow, painstaking process that drew heavily on analytical and medical chemistry skills. Several important classes of drug were discovered this way, including the most important diuretics, many of the most widely used psychoactive drugs and several powerful antibiotics. The chemists working within this regime codified little of the knowledge acquired, so new compounds design was driven by the skills of individual chemist. However, the successful introduction of a new chemical entity has to be considered as quite rare event. From the mid 1970s substantial advantages in physiology, pharmacology, enzymology and cell biology led to enormous progress in the ability to understand the mechanism of action of some existing drugs and the biochemical and molecular roots of many diseases. This new knowledge has a profound impact on the process of discovery of new drugs offering to the researches a significantly more effective way to screen compounds. Moreover, the availability of drugs whose mechanisms of action are well known made possible significant advances in the medical understanding of a number of key diseases. This understanding led to the development of the technique of 'rational drug design'. Researches are now beginning to be able to 'design' compounds that might have particular therapeutic effects. The advent of 'biotechnology' had a significant impact both on organizational competencies required to be a successful player in the pharmaceutical industry and industry structure in general.

After this brief introduction, it could be useful to fix some of the qualitative characteristic about mechanisms and factors affecting industry evolution, stylized in the History-Friendly model. First of all innovative new drugs arrive quite rarely but after the arrival they experience extremely high rates of market growth. This entails a highly skewed distribution of the returns on innovation and of product market size as well as of the intra-firm distribution of sales across products. So, few 'blockbusters' dominate the product range of all mayor firms (Matraves C,(1999)). The industry has been characterized by a significant heterogeneity in terms of firms' strategic orientations, indeed other firms not specialized in R&D and innovation, but in imitation and marketing are able to survive. In general terms the 'oligopolistic core' of the industry as been composed by a stable group of firms, which maintained over time an innovation-oriented strategy. At the same time the industry was characterized by quite low level of concentration both at aggregate level and in the individual sub-markets like e.g. cardiovascular, diuretics, tranquilizers, etc. There is evidence that institutional factors seems to have played a decisive role in the development of pharmaceutical industry. The institutional arrangements surrounding the public support of basic research, intellectual property protection, procedures of product testing and approval, pricing and reimbursement policies have all strongly influenced both the process of innovation and the economic returns (and thus incentives) for undertaking such innovation. Something more may be said about intellectual property protection. Pharmaceutical has historically been one of the few industry were patent provide a solid protection against imitations, for two main reasons. The first is that small variants in the molecule's structure can drastically

alter its pharmacological properties. The second reason is that other firms might undertake research in the same therapeutic class as an innovator, but the probability of finding another compound with the same therapeutic properties that does not infrange on the original patent could be quite small. The procedures of product approval are also very important. Pharmaceuticals are regulated products. Procedures for approval have a deep impact on both the costs of innovating and on firms ability to sustain market position once their product has been approved. Since the early 1960s most countries have steadily increased the stringency of their product approval precesses. However it was the USA, with the Kefauver-Harris Amendament Act in 1962, and the UK, with Medicine Act in 1971, that look by far the most stringent stances among industrialized countries. Federal and Drug Administration (FDA) shifted from a role of essentially an evaluator of evidence and research findings at the end of research process, to an active participant of the process itself. The resources necessary to obtain approval of a new drug application have been largely increased. They probably caused a sharp increase in both R&D costs and gestation times for new chemical entities. Although the process of development and approval increased costs, it significantly increased the barrier to imitation, even when the patent expired. The introduction of tougher regulatory environment in the UK and in USA was followed by a sharp fall of the number of new drug lanced and many small or weak firms exited the market.

4.2 Model setup

In this section we describe the basic features of the model implemented for the analysis of the 'random screening' era (Malerba, Orsenigo (2001) (2002)). This description considers the essence of the model and leaves to the paragraphs 4.2.1, 4.2.2, 4.2.3 and 4.2.4 the specific details. At the beginning, in the age of random screening, a number of firms enter the market and start to interact in the simulation environment. The firms start to invest in random searches for promising molecules that might be the basis for development of a drug in particular therapeutic category. Some molecules seem to be promising for the drug development phase and are patented. After the selection of the molecule that seems more profitable, firms consume time and financial resources to turn it in a marketable drug. The investments in research and development are arranged in fixed share of budget and expose the firms to the risk of running out of money and failure. After the development phase successfully ended, firms engage in marketing activities and begin selling their drugs. Sales are influenced by drug quality, price charged and firms' marketing efforts. At the beginning successful drugs in a particular therapeutic category face no competition. But after some times other firms may discover and develop competing drug. Moreover, after patent expiration, imitation occurs and the market share and revenue of original innovator start to be eroded. For firms the imitation of drugs is less expensive and less time consuming but also less profitable. After a firm has successfully developed a

drug, it begins searching again for a new promising molecule and the process begins all over again. The firms engage sequentially in different projects until the end of simulation or the failure, they will progressively diversify into new therapeutic categories. In the next section we will provide a more detailed description of the simulation model.

4.2.1 Topography

The environment where firms act is composed by several therapeutic categories (TC). Each therapeutic category has a different number of potential customers (patients). The economic size of a TC depends on the number of patients buying and on the prices of the products, in other words is expressed by total sales. Patients of TCs are divided in a certain number of submarkets ($n.sub$) where a minimum level of quality of the drugs is required to sell (QC_{sub}). Thus, some groups of consumer do not buy a drug if its quality is not satisfactory. The number of patients in the overall therapeutic category (N_{TC}) is exogenously given and grows at a certain rate (GN_{TC}) during each simulation period. This is known by the firms. In the model there are n ($n = 100$) therapeutic categories each with a specific number of patients drawn from a normal distribution $N_{TC}\tilde{}N(\mu_{N_{TC}}, \sigma_{N_{TC}})$. Within a therapeutic category there are a certain number of molecules M ($M = 150$), which firms aim to discover and which are the basis of pharmaceutical products. Each molecule is characterized by a certain value of quality Q. A large percentage of molecules have null value of quality ($Qnull$), thus are useless for the firms. Others, from which the firms will generate drugs, have a positive value of quality drawn from a normal distribution $Q\tilde{}N(\mu_Q, \sigma_Q)$. When a molecule is discovered by a firm, the patent protection starts. This grants the protection in two ways. First, the width (PW), prevents competitors from developing all the molecules located in the neighborhood of the patented. Second, specifies the temporal duration of the patent (PD) up to that time in which the molecule become imitable by other firms. Once the drug development is successful, it gets an economic value (PQ) equal to the value of the molecule quality Q_i, where $i = 1, 2, .., 150$ for each TC. The economic value of product influences the demand function for such product, described in section 4.2.3.

4.2.2 The firms

Basic features of firms. At the beginning of the simulation a number F of firms potentially could enter the market. They starts with a budget B ($B = 3000$) equal for all. In the model, firms have a limited understanding of the environment and their behavior follows some simple rules-of-thumb and routines. In particular the firms are engaged in three sequential activities repeated until the exit of the market or the end of simulation: *search, research and marketing*. The first process invests a given share of budget (B_S) in the activity of looking for the promising molecules in the environment. The second process invests another share of budget (B_R) in the activity of developing the molecule in a marketable drug, this is very time and resources consuming

and firms risk to fail running out of money. The residual share of budget (B_M) is invested in marketing activities to promote the selling of the new drug to win the competition of other products in the same TC. Firms are characterized by different 'strategies' and have different propensities of investments in the three activities. At the beginning of the simulation only half of the firms try to enter the market. This group is labelled 'innovators'. After patent expiration, when the first patented molecule become available for all, another group of firms try to enter the market developing products from unpatented molecules. This is the 'free-riding' behavior of the 'imitative' firms. In the model, the marketing propensity of the firms, ϕ,is fixed for the innovators (ϕ_{inno}) and imitators (ϕ_{imi}) and account the relation (ϕ_{inno}) > (ϕ_{imi}). The R&D investments propensity of the firms is then defined by a share of budget $(1-\phi)$ that complements the share of the propensity to marketing investments. Thus, the firms' budget is divided among search, research and marketing activities as follows:

Resources for search (B_S): $(1-\phi)\omega B$

Resources for research (B_R): $(1-\phi)(1-\omega)B$

Resources for marketing (B_M): ϕB

Where ω (randomly drawn from a uniform distribution) is invariant and firm specific.

Innovative activities. In the age of 'random screening' the innovators look for new molecules through the search process. The amount of money invested in search activities, B_S, determines a number (X_{TC}) of TC explored by a firm during its current project, equation (11).

$$X_{TC} = \frac{B_S}{DC} \tag{11}$$

The parameter DC represent the cost of draw of a molecule from a TC. Firm selects randomly some TCs and molecules making X_{TC} extractions. A greater expenditure in search process allows the firms to have more chances to make successful extractions. Moreover there is a fixed cost (FC) that firms have to pay for each period when they are involved in the search activity. Firms have few information about the molecules drawn, due to the limited understanding of the environment, and know only whether Q_i is greater than zero or not. In the case of non-zero quality molecule, and if it has not been patented by others, then patent protection for that molecule is obtained. If a firm experiences successful extractions in more than one TC and thus finds more than one molecule having a positive Q, it chooses to start research activity on the TC which shows the best expected profits. In the case of the absence of products in TC the firm suppose to become leader in the therapeutic category, and calculates expected profits. This criterion addresses the firms to the TC with the greatest number of patients

and less crowed. The patented molecules that are not selected become part of a portfolio of 'sleeping molecules' that can be exploited in every further round of search. If search is successful, the firms move to next activity: the research. Only after a product has been developed or in case the project has failed the firm will start another search iteration.

Imitative activities. For imitative firms the phase of search explanation is quite easy, they do not spent resources to draw a TC. The imitative firms look for already discovered and developed molecules whose patent has expired. When the available molecules with expired patent are in more than one TC, they select the TC with the best profit expected using the same criterion of innovators. In the TC selected for imitation, firms chose the molecule with highest 'perceived' quality, R. R is a function of Q as shown in equation (12).

$$R_i = (1 + \widetilde{\beta})Q_i \tag{12}$$

Where $i = 1, 2, .., 150$ for each therapeutic category and $\widetilde{\beta}$ is drawn from a uniform distribution ($\widetilde{\beta} \in U[-0.1, +0.1]$). Hence, high quality molecules will be more frequently picked up by imitators.

Research activities. Research activity means the actuation of a product development project that 'transforms' a molecule into a drug that can be sold with certain characteristic of quality. This process is very time and money expensive, in particular for innovative research, while is quite cheap for imitative research. Both innovators and imitators do research. A firm, starting with a research budget Br, progresses toward the full development of the drug. That is to say, firm have to 'climb' a fixed number of steps ($s_{total} = 30$) in order to develop a drug having a quality Q. Each steps implies unitary cost CS. Thus, the total cost of developing a drug is equal to $CS \cdot s_{total}$. Firms differ in costs and research investment but have the same number of step to 'climb'. It is clear that the research speed is proportional to the value of B_r and to the research costs. Firms move faster if they pay more in each period according to the following relationship:

$$s_t - s_{t-1} = \alpha \cdot \frac{B_r}{CS}$$

Where $s_t - s_{t-1}$ are the number of steps 'climbed' between t and $t-1$ periods, B_r is the budget spent by the firm in research, CS is the cost of a single step of research, α is a fixed coefficient that proportions budget resources dedicated to research in a single period. The α coefficient and costs CS are different for innovators and imitators.

The costs of a single step of research grows each period at a certain rate CG, starting from a high value for innovation (C_{inno}) and low value for imitation (C_{imi}). The growth represent the increment of expenditure caused by more stringent rules, growing complexity of clinical trials fixed by external agencies (e.g. the FDA) (Grabowsky, H.(2002)). Procedures of product approval of an innovative product are many time more expensive

than the imitative one, hundred millions against few millions of dollars, in particular after the Hatch-Waxman Act in 1984.

Given its budget B_r a firm may be able to buy all the steps to develop a drug. Before commercialization the last obstacle is to warrant a minimum level of quality, requirement to enter the market. There is an exogenous threshold on quality of the product fixed to represent a kind of'quality check' (QC), imposed, for example in the USA, by the Federal and Drug Administration. Below this value the drug cannot be commercialized and the project fails. Reached the minimum quality, the product is labelled as imitation if risen from a molecule with expired patent or otherwise it is labelled as innovative.

Marketing activities. When the develop process come to an end and the quality of the drug is over the threshold for entering the market the firm invests the budget B_M, sets for marketing. The launch of the product time T_L is the moment when the firm creates the 'product image' A_{jT_L}, proportional to marketing spending. Moreover, the firm will profit from a marketing expenditure θ from its previous products $k \neq j$. The 'image' is eroded in the course of time at a rate equal to eA in each period. The level of the 'image', A_{jt} in period t is given by:

$$A_{jTL} = B_{M_{TL}} + \theta \cdot \sum_{k \neq j} [A_k]$$

for $t = TL$

$$A_{jt} = A_{jt-1} \cdot (1 - eA)$$

for $t > TL$.

4.2.3 Demand and market share

In this model, each firm sells a specific quantity of drugs in every period and pays a very low cost for manufacturing every single unit k produced. The costumers are homogenous and we do not distinguish between patients and physicians. Decision to buy a specific drug depends on several factors: quality PQ, price P and image level A_j. The quality of the drug decides the number of submarkets ($n.sub$) reached, thus the number of well-disposed patients to buy. Therefore, low quality drugs, probably with many contraindication, will reach few patients even if there are few competitors in the TC. The number of patients in each submarket (N_{sub}) is give by:

$$N_{sub} = \frac{N_{TC}}{n.sub}$$

The share of a drug in a submarket depends by a 'merit' function ($U_{i,t}$). The factors listed previously, determine value of merit by:

$$U_{i,t} = PQ_i^a \cdot \left(\frac{1}{P_{i,t}}\right)^b \cdot A_{i,t}^c \tag{13}$$

Where PQ_i is the economic value of the drug, equal to quality Q_i, $P_{i,t}$ is the price applied by the firm (see below) $A_{i,t}$ is the image of the product. a, b and c are specific to each TC and drawn from uniform distributions. The market share of product i in the submarket sub of TC is proportional to its relative merit as compared to the other competing drugs in the same submarket and is given by:

$$S_{i,sub} = \frac{U_i}{U_{sub}}$$

Where U_{sub} is the sum of the merit of all products in the submarket. Firms may have a product accessing different submarkets. Thus, their market share in the TC is given by:

$$S_{i,TC} = \frac{\sum S_{i,sub}}{n.sub}$$

When the share of patients for each product is given, it's time for the firms to adjust the selling prices for the next period $(t+1)$ if some market conditions have changed. The price of product i at time t depends, indirectly through the mark-up (mup_i), from its market shares at $t-1$. The prices are set using the two following equations:

$$P_i = k \cdot (1 + mup_i)$$

$$mup_{i,t} = (1-\lambda) \cdot mup_{i,t-1} + \lambda \cdot \left(\frac{S_{i,TC}}{\nu - S_{i,TC}} \right)$$

Where the cost of production is a constant k for all firms, λ is a constant to weight the past mark-up in respect of present share state and ν is an elasticity. Mark-up (mup) is the desire rate of return that each firm wants to obtain from selling its drug. Thus, higher the mup higher the price, but also lower the demand (equation (13)). Moreover the erosion of the market share in a TC by the competition of other products produce an adjustment of prices, on the contrary the monopolistic position of a drug in a TC produces a rise of prices.

4.2.4 Budget accounting and exit rules

Revenue of firm f for product i is $\pi_{f,i}$, and it is given by:

$$\pi_{i,f} = \sum_{1}^{n.sub} [P_i \cdot (S_{i,sub} \cdot N_{sub}) - k \cdot (S_{i,sub} \cdot N_{sub})]$$

$$\pi_{i,f} = \sum_{1}^{n.sub} [K \cdot mup_i \cdot (S_{i,sub} \cdot N_{sub})]$$

because firm f may have more than one product, total revenue (Π_f^{TOT}) is the sum of revenues obtained from all the products of the firm:

$$\Pi_f^{TOT} = \sum_{i=1}^{products_f} [\pi_{i,f}]$$

The revenues, in each period, accumulate in an account that is used as a budget to finance search, research and marketing investments. Thus, we are assuming that firms reinvest their whole budget, without paying dividends to shareholders. Before defining the exit rules it is important to calculate the total share of pharmaceutical market owned by a single firm. It is given by:

$$S_{tot,f} = \frac{\Pi_f^{TOT}}{\sum[\Pi_f^{TOT}]}$$

Firms with very low level of total market share and with a low level of efficiency in the search and research activities exit the market. The three exit rules are given by:

$$Ef > (1-r) \cdot \left(\frac{1}{F}\right) + r \cdot S_{tot,,f}$$

$$\Pi_f^{TOT} < \psi_1 \cdot DC$$

$$\Pi_f^{TOT} < \psi_2 \cdot CS$$

Where r, Ef, ψ_1 and ψ_2 are constants. When one of the rule is satisfied the firm exits the market.

4.3 Parameters of interest

In this section we explore which parameters, in the input and output lists, seem to be promising for the SA. We will test the sensitivity of the model to five main groups of features: *research and development costs, environment and opportunity conditions, regulatory regime, demand and firm's features.* In table (2) we report the input parameters used in the SCM. The meaning of the input parameters is fully explained in section 4.2 but for clearness we report in the table a brief explanation. For some output parameters, reported in table (3), are useful some detailed notes. The first output of the SCM model analyzed is the Herfindahl index, equation (14), which measures the level of concentration in the whole industry (TH).

$$TH = \sum_{i=1}^{F} [S_{tot,i}]^2 \tag{14}$$

This index summarize in a single value the structural features of the market. Another index, calculated almost in the same way, is the mean Herfindahl index (HTC), the only difference is that we consider each TC as a separate market, and in the end we calculate the mean of the indexes of the overall TCs. The mean effective patent life (EPL) represent the average duration

of patent after the drug commercialization. The total number of TC viewed (TCv) and the mean number of TCs viewed by a single firm (TCD) represent quite different concepts. The first index indicates how many TCs where explored by the firms at the end of a simulation (i. e. how many disease classes have a treatment). The latter represent the diversification tendency of the firms. As we explained in section 4.2.2 the firms spend a share of budget for the commercialization of products to improve the value of their 'image'. The ADV index represent the mean level of the product image at the end of the simulation. The last output which needs a brief explanation is the frequency of blockbuster products (BP). These are drugs which maintain a large market share for a long time giving high profits to the firm. We report numerous index to examine each aspect of the SCM behavior. Among the outputs may be identified four big fields of study: *innovation processes, imitation processes, prices and industry structure*. The values of outputs represent the model in the last period of simulation and they can be characterized by continuous or discrete values.

Groups	Inputs	Brief explanation
R&D costs	DC	cost of a single draw of a molecule from a TC
	FC	fixed cost of search activity
	C_{inno}	starting cost of research processes for innovative firms
	C_{imi}	starting cost of research processes for imitative firms
	GC	growth of costs for each period
Environment	Qnull	percentage of molecules with null quality value
	F	number of possible entrants
Reg. regime	PD	patent duration
	QC	quality check
	PW	patent width
Demand	$\mu_{N_{TC}}$	starting mean number of patients in a TC
	GN_{TC}	growth of the number of patients for each period
	θ	share of image advantage from other products
	eA	erosion of product image level for each period
Firm's fe.	ω	shares of search and research activity from R&D budget

Table 2: List of inputs

5 Computer experiments

5.1 Looking for input feasibility

Let consider the vector $x = (x_1, ..., x_k)$ the input of the model. Each input parameter x_i is drawn out from a uniform distribution with a given mean μ_i and a given range $\mu_i \pm 50\%$[1]. Each parameter of the input may be discrete

[1] except for envitonmental opportunity parameter ($Qnull$), quality check (QC), number of possible entrants (F) and growth of number of patients (GN_{TC}) . They are drawn always from uniform distributions but with different ranges.

Groups	Outputs	Brief explanation
Structure	TH	total Herfindahl index
	HTC	mean Herfindahl index in the TCs
	Af_{in}	alive innovative firms
	AF_{im}	alive imitative firms
	Sz_{in}	mean size of innovative firms
	Sz_{im}	mean size of imitative firms
Innovation pr.	N_{in}	number of innovative drugs
Imitation pr.	N_{im}	number of imitative drugs
Prices	P_{in}	mean price of innovative drugs
	P_{im}	mean price of imitative drugs
	Ptot	mean total prices
Others	TCv	number of TC explored by the firms
	TCD	mean number of TC explored by a firm
	fSH	innovative firms share
	BP	blockbuster frequency
	EPL	mean effective patent life
	ADV	mean image level of products

Table 3: List of outputs

or continuos. The mean value μ_i, rises from the i-th component of $\widehat{x} = (\widehat{x}_1, ..., \widehat{x}_k)$ vector considered as the *standard set*. We define the extended output, originated with the input set x, as D_z. Now we introduce the concept of input feasibility set explained in section 2.2. We choose to exclude the less pleasant outputs from D_z using the simple criterion of do not consider less significant results. The first step is to define an output feasibility set F_z through the use of simple conditions on D_z. Thus, we define feasible output all the model results that have the minimum features that allow to define the existence of the pharmaceutical market. Two necessary requirements appear clear. The first is the necessity of existence, at least, of one imitative firm and one innovative firm which compete in the market. The condition of an empty market is not very interesting simply because nothing happens. Moreover, an analysis of the parameters in the state of an empty market may be self-defeating because, under a minimum threshold of difficult conditions, the result is always an empty market. The feasible second state is the condition of low value of concentration. It is widely recognized and is also an empirical evidence that the structure of pharmaceutical industry is characterized by a low level of concentration and our attempt is to remain quite close to the 'History-Friendly' results. Thus, we fix a maximum level of concentration that appears feasible. Once defined F_z the next step is to apply equation (2) to extended output set D_z and then deduce F_x. Imposing input feasibility, causes the loss of input independence as explained in section 2.2. In table (4) we report the input feasibility set F_x (outcome of the process).

From an analysis of the F_z and D_z (not reported here for brevity) emerges that the independence of the inputs is quite preserved, only few results are

Inputs	Min	Max	Mean	Std
DC	10.0049	29.9854	19.2045	5.8615
FC	50.1516	149.7363	98.1043	29.0152
C_{inno}	10.0161	29.9749	18.3594	5.3455
C_{imi}	0.5009	1.4995	1.0016	0.2932
GC	0.005	0.015	0.0098	0.0029
Qnull	0.9	0.94	0.9199	0.0116
F	25	99	65.1651	20.7166
PD	10	29	20.3601	5.5308
QC	30	44	36.2445	4.1495
PW	0	9	4.3097	2.8663
$\mu_{N_{TC}}$	250.3094	749.9139	524.4019	137.5147
GN_{TC}	0	0.002	0.001	0.0006
θ	0.05	0.15	0.0995	0.0287
eA	0.005	0.015	0.0099	0.0029
ω	0.05	0.15	0.101	0.0282

Table 4: F_x

out of the feasible set imposed as threshold and the input parameters x have weak dependence structures. Thus, as approximation, the input distribution may be considered with independent components.

5.2 Simulations

The extended output set z is composed by 30000 simulations. Each z represent a single 'story' of the industry told by the model under the conditions determined by the inputs. For a given set of inputs $x = (x_1, ..., x_k)$ the first two conditional moments of the distribution $P(z|x)$ are given by equation (15). The first moment values ($\mu(x)$) inside output feasibility set, F_z, are $m(= 1272)$. For a given x there are $n(= 20)$ outputs , thus the total number of outputs generated by the model under the condition of input feasibility set are $m \cdot n(= 25440)$ values of z.

$$\begin{aligned} \mu(x) &= E(z|x) \\ \tau^2(x) &= Var(z|x) \end{aligned} \tag{15}$$

5.3 Multivariate structure I/O

The output $z = (z_1, ..., z_r)$ of SMC is multivariate. The large part of outputs z_j are characterized by strong correlation. We define the variance-covariance matrix as $V_z = Var(z)$ (not reported for brevity). The matrix V_z shows frequent high values of correlation among the outputs. This result is not surprising if we think about the output structure. We describe now one example of positive strong correlation. It concerns the number of innovative alive firms (AF_{In}) at the end of simulation and the number of innovative

drugs sold on the market (N_{In}). It is quite intuitive that the two outputs are strongly correlated, considering that the only sources of wealth for the firms are the sold products.

As we said in section 2 we are able to separate the quota of the global output uncertainty. This depends on the stochastic game ξ and may be related to the inputs variability using ANOVA decomposition of variance, equation (1).

Outputs	Total unc.	Input unc.	Stochastic game unc.	η^2
TH	0.01516	0.003322	0.01184	0.219
HTC	0.007099	0.004406	0.002697	0.6202
AF_{in}	10.05	6.831	3.228	0.679
AF_{im}	104.5	59.19	45.32	0.5662
Sz_{in}	4.86E+07	2.25E+07	2.61E+07	0.4627
Sz_{im}	1.13E+07	3.35E+06	7.94E+06	0.2967
N_{in}	1383	1110	273	0.8026
N_{im}	2722	1765	957.9	0.6481
P_{in}	0.05746	0.01289	0.04458	0.2242
P_{im}	0.01622	0.004062	0.01216	0.2504
P_{tot}	0.008699	0.004823	0.00388	0.554
TCv	530.7	425.9	105.2	0.8019
TCD	396.4	308.7	87.88	0.7783
fSH	0.06551	0.03405	0.03149	0.5193
BP	8.512	5.619	2.897	0.6596
EPL	28.19	7.723	20.47	0.2738
ADV	5.52E+06	3.67E+06	1.85E+06	0.6644

Table 5: Stochastic game and sistematic componets

In table (5), η^2 is the accountable quota of output uncertainty that may be related to each single input using SA techniques. We decide to summit a particular analysis to the outputs with a significative accountable uncertainty, i.e. with a small component of stochastic game. As we said, we suppose that the influence of the input parameters and stochastic game may be separated and the influence of x may be modelled by an appropriate emulator that we investigate. This criterion allows us to accept for the analysis few output parameters. In particular, arbitrarily fixed a minimum threshold of access, we choose only three outputs. These are the number of TCs explored (TCv), the differentiation across TCs of the firms (TCD) and the number of innovative products (N_{In}). The outputs selection evidences the strong dependence from stochastic game of the SCM, thus from the generation of uncontrolled uncertainty inside it. Further analysis including more sources of uncertainty of the SCM may reduce the component of the stochastic game. Moreover, we note from the variance-covariance matrix V_z that two of the three outputs selected are characterized by a strong correlation. In fact, emerges that the number of innovative products (N_{In}) and the therapeutic category explored by the firms (TCv) have a positive correlation $V_z(z_{N_{In}}, z_{TCv}) = 0.96$. Not sur-

prisingly the random draw mechanism on the environment ('random search') produces a spread placement of innovative products in the TCs, thus many products means more probability to explore different TCs. Follows that we decided to investigate only the emulator components of the number of innovative products (N_{In}) considering that the behavior of TCv is quite the same.

5.4 Computer simulation results and data analysis

In this section we examine the input effect only. As explained in section 3.3.1, for a preliminary estimate of coefficients β to perform the SA we can consider the average values of outputs $\widehat{\mu}_i$, $i = 1, ..., m$ and apply ordinary least squares to data $(\widehat{\mu}_i, x^i)$. After, we proceed eliminating the unimportant parameters and make the graphical analysis of the residual. Following this lines we searched for a statistical model having residuals almost independent from the x and normally distributed or at last symmetric around zero, with small Mean Squared Error (MSE). Moreover we searched for a little complexity as measurer by Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Thus, the first step is to eliminate unimportant parameters with t tests. To identify a threshold, common to all the model β coefficients, for the p-values we exploit the optimization of the model complexity in accordance with AIC and BIC. To do this we consider the whole multivariate output and we calculate the variance of the residuals as explained in equation (16).

$$\widehat{\sigma}_{\varepsilon}^2 = Var(V_{\varepsilon} \cdot \alpha') \tag{16}$$

Where V_{ε} matrix contains the residuals of the regressions and $\widehat{\sigma}_{\varepsilon}^2$ is the variance of the linear combination $V_{\varepsilon} \cdot \alpha'$. Moreover, we consider k the whole number of parameters in the multivariate model and m as the number of observations $\widehat{\mu}_i$. AIC and BIC, functions of $\widehat{\sigma}_{\varepsilon}^2$, k and m, are both optimized for a threshold p-value of 10^{-7}.

After the model simplification we focus on a deep analysis of few outputs, selected, as explained in section 5.3, for the low component of stochastic game. At the end of this Section, in table (6) and table (7), where are listed the importances of the input parameters, we decided to report all the SIs in order to give an overall description of the model through the univariate and multivariate SAs even if the stochastic game component largely determinates some outputs. For example, we observe that the number of patients in the TCs influences for the 10.29% the total Herfindahl index, being the maximum quota of accountable output uncertainty only the 21.9%.

5.4.1 Number of innovations (N_{In})

Now we explore the number of innovations in the market (N_{In}) because it has a high value of η^2 and so variance depends largely from inputs. First of all we simplify the model omitting the unimportant parameters and we get a

simple linear model for N_{In} (equation (17) with standard deviation reported in brackets for all the coefficients) with a quite good fitting $R^2 = 81\%$. From figure (1) and figure (2) we search for some non linearities but the residuals seems to be quite well distributed.

$$\begin{aligned} N_{In} = &-0.25(\pm 0.0122)DC + 0.30(\pm 0.0122)F - 0.64(\pm 0.0127)C_{inno} \\ &+ 0.39(\pm 0.0128)PD - 0.56(\pm 0.0123)QC - 0.14(\pm 0.0122)GC \\ &+ 0.40(\pm 0.0125)\mu_{N_{TC}} - 0.09(\pm 0.0122)FC + e \end{aligned} \quad (17)$$

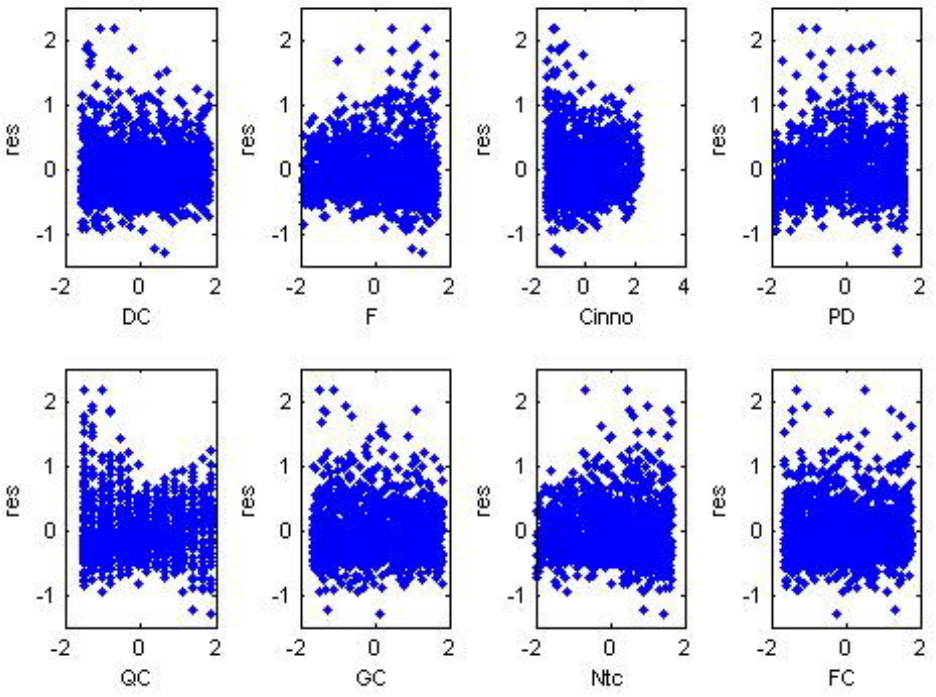

Figure 1: N_{In} Residuals vs. input parameters

About the frequencies of the residual we note high skewness ($s_k = 1.05$) and Kurtosis ($k = 5.42$) but we let analysis of superior order for future studies and, as we previously said we consider only the linear components. Only as a hint we report in figure (3) the residual frequencies considering the quadratics and interactions components. The fitting becomes good with value of $R^2 = 95\%$, but the linear components change value lacking the orthogonality characteristic of the inputs. Thus, for a preliminary estimation, we prefer to maintain the model more parsimonious in the number of estimated coefficients.

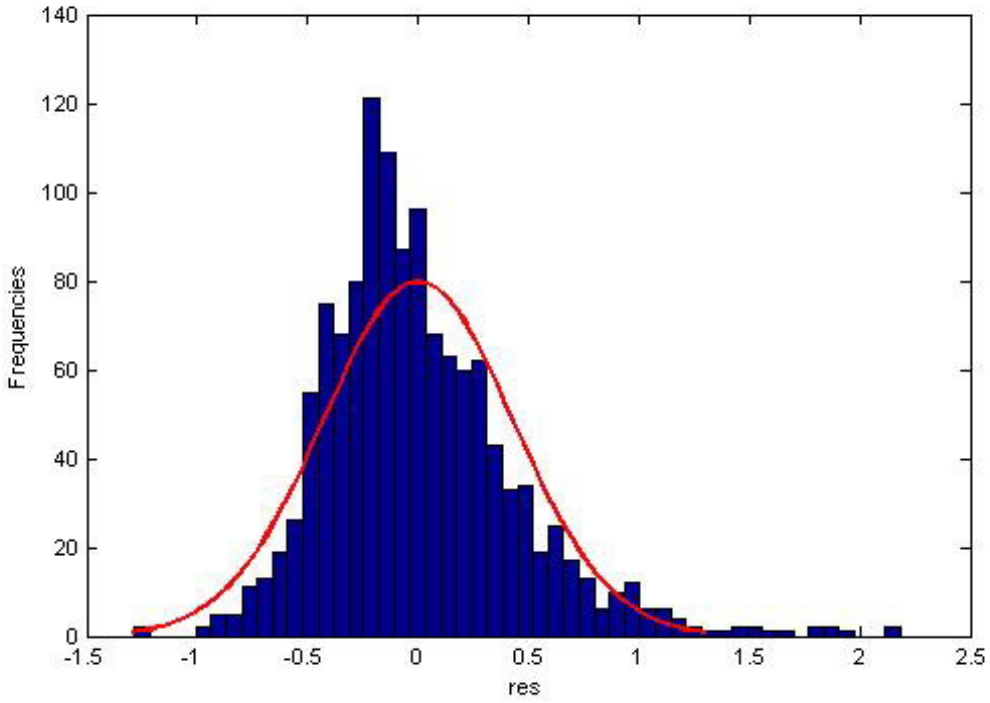

Figure 2: N_{In} Frequencies of the residuals

5.4.2 Diversification of the firms in the TCs (TCD)

The second analyzed output is the mean number of the TCs explored by a firm at the end of simulation (TCD) (model shown in equation (18)). We concentrate on this output because it has a good η^2 and thus has low dependence with the stochastic game. Unfortunately the fitting of the model is quite low with $R^2 = 55\%$ and with high skewness and Kurtosis. The residuals presented in figure (4) and figure (5) are quite normally distributed excepted for the number of possible entrants (F). Probably some non linearities and heteroskedasticity determinates the residual behavior. We try to add quadratic and linear components but the fitting remains low, $R^2 = 69\%$. we leave to future analysis an explanation of the relation between F and residuals.

$$TCD = -0.16(\pm 0.0192)DC - 0.34(\pm 0.0193)F - 0.42(\pm 0.0194)C_{Inno} - 0.43(\pm 0.0193)QC + 0.24(\pm 0.0194)\mu_{N_{TC}} + e \quad (18)$$

5.4.3 Results and discussions

Tables (6) and (7) summarizes SA, including SIs from both univariate models and multivariate ones. The last column reports the multivariate SIs of equation (10).

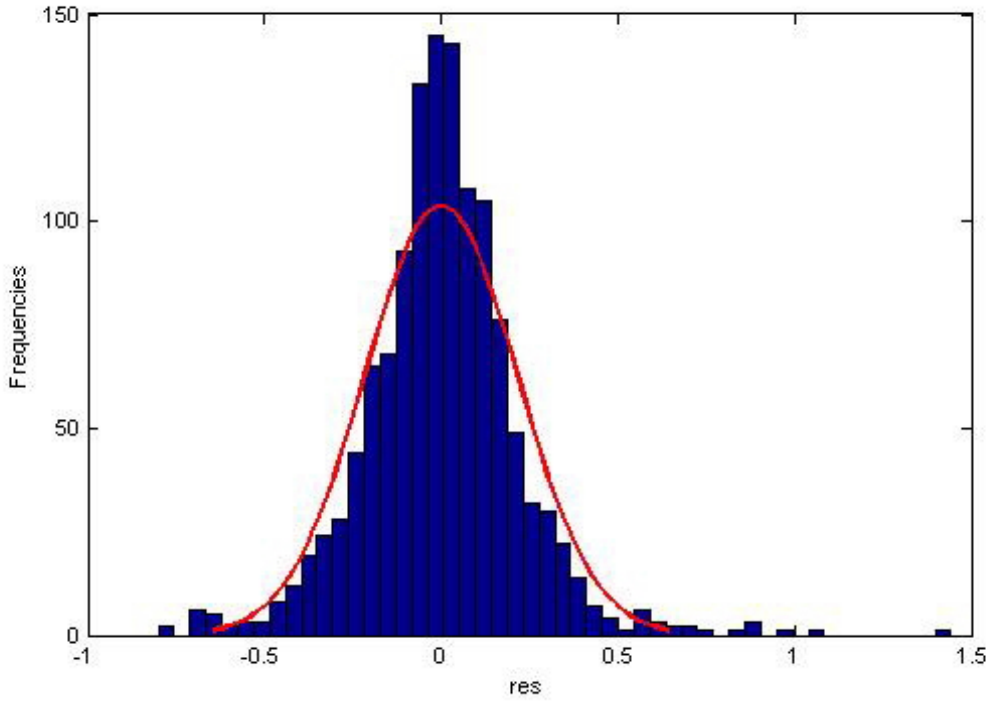

Figure 3: N_{In} Residuals of the model with quadratic and interaction components

	TH	HTC	AF_{in}	AF_{im}	Sz_{in}	Sz_{im}	N_{in}	N_{im}	P_{in}
DC	1.92%	2.73%	6.23%	3.79%	1.13%		5.04%	3.80%	3.42%
FC		0.71%	1.89%		0.24%		0.71%		
C_{inno}		23.86%	37.06%	24.72%	7.16%	6.46%	34.05%	31.96%	32.24%
C_{imi}									
GC	2.09%		0.91%				1.58%		0.91%
Qnull		1.69%			1.19%				0.88%
F		43.31%	8.28%	22.50%	7.42%	84.09%	7.73%	7.52%	17.34%
PD	71.29%	25.16%	13.83%	19.34%	43.12%		12.43%		31.51%
QC	14.21%	1.13%	22.17%	12.44%	0.90%		25.39%	33.93%	10.07%
$\mu_{N_{TC}}$	10.50%	1.40%	9.17%	17.22%	38.06%	9.45%	13.07%	22.79%	3.62%
GN_{TC}					0.01%				
θ									
eA			0.46%						
R^2	0.55	0.79	0.83	0.61	0.78	0.48	0.81	0.49	0.67
η^2	0.22	0.62	0.68	0.57	0.46	0.30	0.80	0.65	0.22

Table 6: SA 1/2

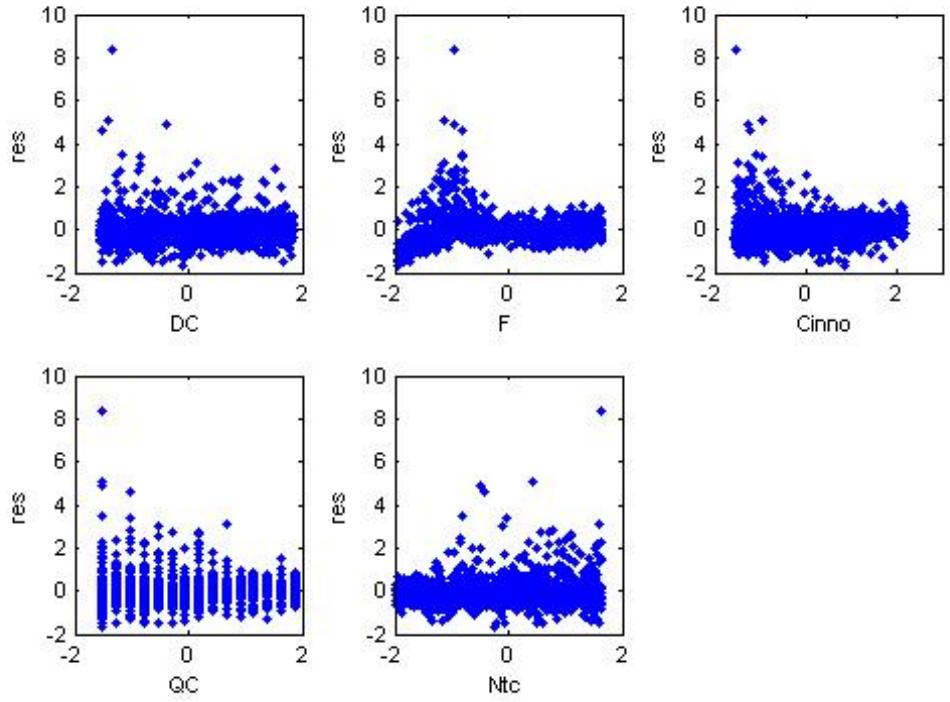

Figure 4: TCD Residuals vs parameters

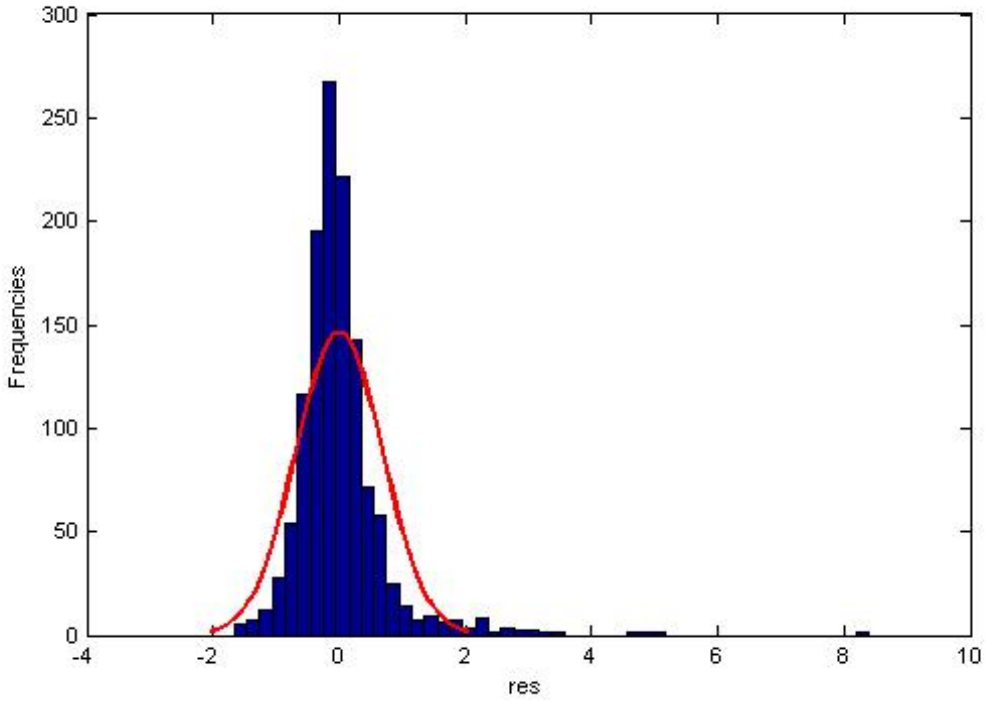

Figure 5: TCD Residuals frequencies

	P_{im}	P_{tot}	TCv	TCD	fSH	BP	EPL	ADV	Multivar.
DC	2.09%	2.39%	4.14%	4.56%	1.86%	3.36%	14.77%	2.44%	4.59%
FC			0.82%			0.66%			0.24%
C_{inno}	20.41%	23.28%	35.42%	31.49%	24.36%	34.74%	54.70%	16.72%	28.54%
C_{imi}					0.32%				0.00%
GC			1.20%		2.57%	0.62%	6.80%		0.36%
Qnull						1.07%	7.02%		0.05%
F	62.97%	47.98%	7.92%	21.54%		18.57%	6.93%	26.01%	16.85%
PD	11.42%	21.80%	9.97%		45.61%	27.29%	2.16%	14.76%	25.69%
QC	3.12%	4.55%	29.78%	32.39%	10.45%	7.41%	3.74%	6.00%	11.69%
$\mu_{N_{TC}}$			10.76%	10.03%	13.62%	6.28%	3.88%	20.29%	11.78%
GN_{TC}									0.00%
θ					0.43%			4.86%	0.01%
eA					0.78%			8.92%	0.18%
R^2	0.60	0.69	0.90	0.53	0.87	0.81	0.57	0.61	
η^2	0.25	0.55	0.80	0.78	0.52	0.66	0.27	0.66	

Table 7: SA 2/2

It can be observed that the investigated input parameters do not influence the SCM to the same extent. In particular the multivariate SI in the last column of table (7) shows the cost of innovative research (C_{inno}), being the 28.54% its overall importance. The second parameter characterized by great influence on the model outputs is the patent duration (PD), scoring the 25.69% of influence. These conclusions agree with many economic works that widely recognize the importance of development costs and regulatory regimes as determinant features of the industry. In particular, considering the univariate SAs of the statistical model, the costs of innovation (C_{inno}) are the decisive parameter for the amount of both alive innovative (AF_{in}) and imitative (AF_{im}) firms, explored TCs (TCv), number (N_{In}) and prices (P_{in}) of innovative products, the effective patent duration (EPL) and the frequency of blockbusters (BP). Furthermore the patent duration (PD) is extremely important in the definition of total Herfindahl index (TH), innovative firms' size (Sz_{In}) and market share (fSH). Moreover relevant importance also persists in mean concentration (HTC) in the TCs, prices of innovative products (P_{In}) and in blockbusters frequency (BP). It is important to remember that for some industry characteristics, the stochastic game has a relevant role and so, for example, in total Herfindahl index the weight of patent duration must be reconsidered. The third most important parameter in Multivariate SA is the number of potential entrants (F) with 16.85% of influence. The following three parameters in order of importance are quality check (QC), the mean number of patients ($\mu_{N_{TC}}$) and the cost of search (DC). The other parameters seem to have a less relevant role in the model. The most important classes of parameters seem to be the R&D costs, in particular cost of innovative research, and regulatory regime. Remember that we are not analyzing

the stochastic game component, largely important for many characteristics of the market, but we rank the inputs x according to the influence on output uncertainty of emulator, that is a statistical simplified model. The amount of innovative products (N_{In}) is an interesting case of univariate SA because has low stochastic game component and good fitting. The variance of this output depends closely from input variance and thus the SA on emulator is very meaningful. The most important parameters, considering as output the number of innovative products, are the cost of innovative development (C_{inno}) and the parameters of the regulatory regime ((QC) and (PD)) . Also the number of patients ($\eta_{N_{TC}}$) plays a key role.

6 Conclusions

A statistical procedure has been proposed for preliminary analysis to evaluate the influence of some parameters in a History-Friendly stochastic computer model of the pharmaceutical industry. The parameters analyzed deal with demand, regulatory regime, costs, environmental characteristics and firm's features. At first we assumed that the influence of the input parameters and the stochastic game on the output variance could be separated. Thus, we modelled the influence of the inputs by an appropriate emulator which is a simplified statistical model. After, using the emulator, we carried on ranking the parameters through the univariate and multivariate SAs. In the latter method, the influence of the inputs on the stochastic computer model has been assessed. Considering the entire aspects of the pharmaceutical industry investigated, the most important parameters result to be the innovation costs and the regulatory regime. Considering the number of innovations, as a case of particular univariate SA characterized by low value of the stochastic game, emerges that costs of innovation and the regulatory regime play also a decisive role in determining the productiveness of the innovative process. This path of analysis allows a preliminary evaluation. In some cases inserting heteroskedasticity, quadratic components, interactions and detailed study of the stochastic game component of the mixed linear model could improve the accuracy of the results. The next step (a proposal for future works) will be to extend the analysis of the model, introducing the omitted components. However we underline the importance of this analysis in order to summarize the large and complex problem of understanding and ranking what features play a relevant role in a market model for pharmaceutical industry. Just at this level of widening the procedure gives useful indications about the outputs of the model and the subordinate economic theory.

References

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

Fassò A. (2006*a*) Sensitivity Analysis and Water Quality. *Working Paper GRASPA*, **23**, (www.graspa.org). In print on: Wymer L. Ed, *Recreational Beaches: Statistical Framework for Water Quality Criteria and Monitoring*. Wiley.

Fassò A. (2006*b*) Sensitivity Analysis for Environmental Models and Monitoring Networks. In: Voinov, A., Jakeman, A., Rizzoli, A. (eds). *Proceedings of the iEMSs Third Biennial Meeting: "Summit on Environmental Modelling and Software"*. International Environmental Modelling and Software Society, Burlington, USA, July 2006. CD ROM. www.iemss.org/iemss2006/sessions/all.html

Fassò A., Esposito E., Porcu E., Reverberi A.P., Vegliò F. (2003) Statistical Sensitivity Analysis of Packed Column Reactors for Contaminated Wastewater. *Environmetrics*. **14**(8), 743 - 759.

Fassò A., Perri P.F. Sensitivity (2002) Analysis. In Abdel H. El-Shaarawi and Walter W. Piegorsch (eds) *Encyclopedia of Environmetrics*, **4**, 1968–1982, Wiley.

Grabowski, H. (2002) Patents and new products in the development of pharmaceutical industry

Grabowski H., Vernon J, (1992), Brand loyalty, Entry, and price competition in pharmaceuticals after the 1984 drug act, *Journal of Low and Economics,* Vol.35,N.2,pp.311-350

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

Kennedy M.C. O'Hagan A. (2001) Bayesian calibration of computer models, *J. Royal Stat. Soc. B*, **63**, 425-464.

Lee, Y., Nelder, J.A. (2001) Hierarchical generalized linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, 88, 987-1006.

Malerba F., Nelson R., Orsenigo L., Winter S., (2001),Competition and industrial policies in a History Friendly model of evolution of the computer industry,*International Journal of Industrial Organization*,19,635-664

Malerba F., Nelson R., Orsenigo L., Winter S., (1999),History Friendly models of industry evolution: the computer industry,*Industrial and Corporate Change*, 8, 1

Malerba F., Orsenigo L. (2002) Innovation and market structure in the dynamics of the pharmaceutical industry and biotechnology: towards a history friendly model, *Industrial and Corporte Change*, 11, 667-703

Malerba F., Orsenigo L. (2001), Towards a History Friendly model of innovation, market structure and regulation in the dynamics of the pharmaceutical industry: the age of random screening. WP n.124

Matraves C., (1999) Market structure R&D and advertising in the pharmaceutical industry. *The Journal of Industrial Economics,* XLVII

McKay M. D., Beckman R. J., Conover, W. J. (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21 , 239-245.

Saltelli A., Chan K., Scott M. (2000) *Sensitivity Analysis*, Wiley, New York.

Saltelli A., Tarantola S., Campolongo F., Ratto M. (2004) *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Wiley.