Exploring Greek innovation activities: the adoption of generalized linear models

George Halkos

University of Thessaly, Department of Economics

2010

Exploring Greek Innovation activities: The adoption of Generalized Linear Models

George Halkos[1] and Christos Kitsos[2]

[1] Department of Economics, University of Thessaly

[2] Technological Educational Institute of Athens

Abstract

In this paper we examine the innovative performance of Greek firms in terms of the women participation in research and technological development. For this reason we rely on the final results of a research project on women in innovation, technology and science, based on 279 questionnaires selected on a two years time period (2004-2006). Concerning the female participation in innovations a number of variables are used, like the total number of women employees by age, by education level, by firm size and by sector, as well as women in product and in process innovations, their position in the firm (owner, manager) and finally equality in job enrichment, in salary, in education–training and in promotion. Apart of presenting the empirical results relying on the analysis of the data collected by the survey to the Greek enterprises, we use the collecting variables in an econometric formulation using logistic regression and extracting the associated probabilities for implementing innovations. For this reason, first the General Linear Model (GLIM) is introduced and statistical inference and estimation problems are discussed. Then the Logit Model is presented under the theoretical framework of the Generalized Linear Models (GLIM), while some theoretical inside is extended with a number of suggested propositions and theorems.

Keywords: Innovation, Entrepreneurship, Competitiveness, Diversity.

JEL Codes: O31, Q55, L26, C10

Address for Correspondence

George Emm. Halkos
Director of Postgraduate Studies
Associate Editor in Environment and Development Economics
Director of the Operations Research Laboratory
Deputy Head
Department of Economics
University of Thessaly
Korai 43
Volos 38333, Greece
Tel. 0030 24210 74920
FAX 0030 24210 74772
email: halkos@uth.gr
http://www.halkos.gr/

Introduction

Innovation refers to a new or significantly improved product (good or service) introduced to the market. Innovations rely on the results of new technological developments, new combinations of existing technologies and production methods or the utilisation of other knowledge acquired by the firm during its operation. Specifically product innovation may take place with respect to its fundamental characteristics, its technical specifications, potential uses, or user friendliness. Innovation may also refer to the introduction within a firm of a new or significantly improved process. A process innovation includes new and substantially improved production technology, better and easier methods of supplying services and of delivering products. Innovations may be developed either by the innovating firm or by another firm. Innovations should be new to the firm under consideration. For product innovations they do not necessarily have to be new to the market and for process innovations the firm does not necessarily have to be the first to have introduced the process.

The emphasis in the existing literature is mainly on an increasing relevance of knowledge and innovation as an input to production and innovative processes (OECD, 2001). The OECD (1996) report on 'The Knowledge-based Economy' clarifies the terms used in describing the 'New Economy'. The increasing contribution of high-tech sectors (computers, electronics and aerospace) to GNP and employment as well as the recognition of the role of knowledge and technology in economic growth has led to the establishment of the term 'knowledge-based economy' (OECD 1996, p. 9).

In innovation management, the first explicit theory is the "technology push theory or engineering theory of innovation", where basic research and industrial R&D are the sources of new or improved products and processes. The production and uptake of research follows a linear sequence from the research to the definition of a product and specifications of production. Alternatively and in the 1960s, the "market pull theory of innovation" gave a central role to research as a source of knowledge to develop or improve products and processes.

The latter theory recognizes for the first time the organisational factors as contributors in innovation theory. Technical feasibility was still considered as necessary condition for innovation but no longer sufficient for successful innovation (Schmookler, 1996; Myers and Marquis, 1969). Here is where a new generation called the "chain-link theories" of innovation emerged in order to explain that linkages between knowledge and market are not as automatic as assumed in the technology push and the market pull theories of innovation (Von Hippel, 1994).

At the end of the 1980s and during the 1990s, a technological networks theory of innovation management was developed by a new group of experts as "systems of innovation". According to Nelson (1993) and OECD (1999) this view emphasizes the significance of external to the firm information sources such as clients, suppliers, consultants, government laboratories and agencies etc. Finally, "the social network theory" of innovation management may be considered which states that knowledge is crucial in facilitating innovation. According to Foray (2000) the increasing and steadily growing significance of knowledge as a production factor and as a determinant of innovation is explained by the continuous accumulation of technical knowledge through time, as well as by the use of communications technologies that facilitates this knowledge making it available instantly worldwide.

The target of this paper is to analyse the framework, the obstacles and the determinant factors and furthermore the role of female entrepreneurship in the Greek firms, under the Generalized Linear Models statistical framework. Specifically, we

provide the theoretical background of the adopted Generalized Lineal Model theory, so crucial in any Risk Analysis problem and we also extensively discuss the Greek case of female participation in innovation activities of firms. We are relying on the results of a research project on women in innovation, technology and science, based on 279 questionnaires selected on a two years time period (2004-2006), when 2200 questionnaires were delivered across the country, to firms with more than 20 employees. In this paper cconcerning the female participation in innovations a number of variables are used. Specifically, the total number of women employees by age, by education level, by firm size and by sector, as well as women in product and in process innovations, their position in the firm (owner, manager) and finally equality in job enrichment, in salary, in education–training and in promotion.

Apart of presenting the empirical results relying on the analysis of the data collected by the survey to the Greek enterprises, we use the collected variables in an econometric formulation using logistic regression and extracting the associated probabilities for implementing innovations. For this reason, first the General Linear Model (GLIM) is introduced and statistical inference and estimation problems are discussed. Then the Logit Model is presented under the theoretical framework of the Generalized Linear Models (GLIM), while some theoretical inside is extended.

The structure of the paper is the following. Section 1 introduces the theoretical background of the GLIM while section 2 discusses the logistic model as a special case of GLIM. Section 3 presents an empirical application in the case of innovation activities in Greece. In particular this section discusses the data used, the basic descriptive empirical findings and the results derived using a logistic regression formulation. The final section concludes the paper.

1. Generalized Linear Models (GLIM)

1.1 Linear Models

Models of the general form

$$y = \mathbf{X}\beta + e \qquad (1.1.1)$$

where y is a n x 1 vector of responses, β is a p x 1 vector of parameters, X is a n x p matrix with elements zeros or ones or values of the so-called "independent" or "input" or "explanatory" variables and e is an unobserved n x 1 vector of errors attracted interest the 20th century. There is an extended bibliography in English and in Greek, see Halkos (2006), Kitsos (2006) among others. All the theory of Linear Models (or General Linear Model – GLIM) is based on the hypothesis that the errors are independent and identically distributed with normal distribution $N(0,\sigma^2)$.

But it has been noticed that the responses, y, might have distribution other than the Normal or might be categorical rather than continuous. And even more the "link" between the responses y and the explanatory variables, with form the matrix **X**, might not be linear of the form (1.1.1). Although these are disadvantages there are some advantages to keep the theoretical balance:

1. Many of the properties of the Normal distribution can be faced to a wider class of distributions (see section 1.2)

2. There are numerical extensions from the estimation of the form Xb to the general form g(Xb) (see section 2)

These two main points provided evidence for the development of the Generalized Linear Models theory, covering the General Linear Models.

1.2 Exponential family of distributions

Consider a single random variable Y whose probability function is either discrete or continuous. We shall say that it belongs to the exponential family if the probability function can be written as

$$f(y;\theta) = \gamma(y)\, t(\theta)\, e^{a(y)b(\theta)} \qquad (1.2.1)$$

where a(.), b(.), γ(.), t(.) are known functions, $\theta \in \Theta \subseteq \mathbb{R}^p$ is the vector of unknown parameter from the parameter space $\Theta \subseteq \mathbb{R}^p$. In practice usually p=2, that is only two parameters have to be estimated. In most theoretical approaches Θ is assumed compact so that if there are any sequences assumed (as in sequential analysis problems) the sequences converge within Θ.

The (1.2.1) form can be reduced to

$$f(y;\theta)=\exp\{a(y)b(\theta)+C(\theta)+d(y)\} \qquad (1.2.2)$$

where $\gamma(y)=e^{d(y)}$, t(θ)=exp[C(θ)].

If γ(y)=y the (1.2.2) is known as the *canonical form*. The term b(θ) is known also as the *natural parameter* of the distribution. If there are additional parameters, besides θ, they are acting as *"nuisance parameters"* forming part of a, b, c, d and they are assumed to be known (although might not be!)

For the well-known distributions Poisson, $P_0(\lambda)$, the normal $N(\mu, \sigma^2)$ and the binomial B(n; y, p) which belong to the exponential family. Table 1 below summarizes the mentioned terminology.

Table 1: Poisson (λ), Normal $N(\mu, \sigma^2)$, Binomial(n; y, P) from the exponential family

Distribution	**Natural parameter**	**C**	**d**
$P_0(\lambda)$	Logλ	-λ	-logy!
$N(\mu, \sigma^2)$	$\frac{\mu}{\sigma^2}$	$-\frac{1}{2}\left[\frac{\mu^2}{\sigma^2}+\log 2\pi\sigma^2\right]$	$-\frac{1}{2}\frac{y^2}{\sigma^2}$
B(n; y, P)	$\log\frac{P}{1-P}$	$n\log(1-P)$	$\log\binom{n}{y}$

Notice that from (1.2.2) the likelihood function *L* (the joint distribution function for the independent $Y_1, Y_2, ..., Y_n$ from (1.2.2)) is

$$L=\prod_{i=1}^{n} f(y_i;\theta)=\exp\left\{b(\theta)\sum_{i=1}^{n} a(y_i)+nG(\theta)+\sum_{i=1}^{n} d(y_i)\right\} \qquad (1.2.3)$$

The statistic $T=\sum a(y_i)$ is called *sufficient statistic* for b(θ). Practically this means that the statistic T, the summation, summarizes all the information about θ.

Now, the target is to find "closed", if possible, expressions for the expected value and variance, so that inference to be feasible for a(y). The main implication is the so called Granger-Rao theorem (see Kitsos and Edler, 1997) and for a theoretical framework Scherrish (1995). Let us now discuss how this can be useful in practical problems.

Consider the log-likelihood *l* and $U=\frac{\partial l}{\partial \theta}$, $U'=\frac{\partial^2 l}{\partial\, \theta^2}$. Then it can be proved that

$$E(U)=0, \qquad Var(U)=E(U^2)-E(-U') \qquad (1.2.4)$$

Function U is known as *"score function"* and the variance of U is a case of what is called in Statistics (Fisher's) information. There are different measures of information, like Fisher's, Shannon's etc. and there are some relations among them, but this is beyond the task of this paper. Moreover we would like to emphasize that the closed forms we are looking for, do not provide solutions and therefore an old (and secure) iterative scheme is adopted: Newton-Raphson.

Here we comment that these methods are rich in theoretical background and mathematical application. So the experimentalists, and not only, were wondering how useful are really. Now with the statistical packages the results are easy to be obtained, but to interpret them, the theoretical insight is necessary. That is what we provide, in a compact form, in this section.

Consider the exponential family as in (1.2.2). The log-likelihood is

$$l = \log f(y;\theta)=a(y)b(\theta)+c(\theta)+d(y) \qquad (1.2.5)$$

Therefore we easily evaluate the score function and its derivative:

$$U= a(y)b'(\theta)+c'(\theta), \quad U'= a(y)b''(\theta)+c''(\theta) \qquad (1.2.6)$$

As E(U)=0 then

$$0 = E[a(y)]b'(\theta)+c'(\theta) \Rightarrow E[a(y)]=-\frac{c'(\theta)}{b'(\theta)} \qquad (1.2.7)$$

Thus a closed form for E[α(y)] has been evaluated. Moreover, consider (1.2.4), it is

$$Var(U)=[b'(\theta)]^2\ Var\ [a(y)]$$

$$E(-U') = -b''(\theta)\ E\ [a(y)] - c''(\theta)$$

Thus

$$Var\left[a(y)\right] = \frac{1}{[b'(\theta)]^2}\left[b''(\theta)c'(\theta) - c''(\theta)b'(\theta)\right] = \frac{1}{[b'(\theta)]^2}\begin{vmatrix} c'(\theta) & b'(\theta) \\ c''(\theta) & b''(\theta) \end{vmatrix} \tag{1.2.8}$$

Therefore with (1.2.7) and (1.2.8) *closed forms* for *the expected value and variance* of a(y), from the exponential family of models, have been evaluated.

1.3 Definition of GLIM

Nedler and Wedderburn (1972) in their pioneering paper noticed the unity a class of statistical methods, involving linear combinations of parameters. The idea of a generalized linear model came into light. We briefly follow it.

Consider $Y_1, Y_2, ... Y_n$ independent random variables, each with a distribution form the (1.2.1) exponential family. Moreover:

i. Each of the observations Y_i, i=1,2,...n has distribution of the canonical form i.e. a(y)=y. Depending on a single θ_i (1.2.2) is reduced to

$$f(y_i;\theta_i)=\exp\{y_i b(\theta_i)+C(\theta_i)+d(y_i)\} \tag{1.3.1}$$

Notice that not all θ_i's have to be the same

ii. The joint probability density function of $Y_1, Y_2, ... Y_n$ is then evaluated as

$$f(y_1, y_2, ..., y_n; \theta_1, \theta_2, ..., \theta_n) = \exp\left\{\sum_{i=1}^{n}[y_i b_i(\theta_i) + C(\theta_i) + d(y_i)]\right\} \tag{1.3.2}$$

iii. We consider a "small" set of parameters $\beta_1, \beta_2, ..., \beta_p$ say p<n and a (monotone differentiable) function g, known as *link function* which relates the expected value of $Y_i, E(Y_i) = \mu_L$, with a linear combination of β's, i.e.

$$g(\mu_L) = \mathbf{X}_L^T \beta \tag{1.3.3}$$

with $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$, $\mathbf{X}_i$ and a p x 1 vector of explanatory variables.

Example 1: The general linear model of the form (1.1.1) $y = \mathbf{X}\beta + \mathbf{e}$ is, trivially, a GLIM model with link function $g(\mu_i) = \mu_i = \mathbf{X}_i^T \beta$, i.e. the identity and $Y_i \square N(\mu_L, \sigma^2)$.

Example 2: For the binary responses $Y_1, Y_2, ... Y_n$, with $P(Y_i = 1) = P_i = 1 - P(Y_i = 0)$. The probability function of Y_i is $p_i^{y_L}(1-p_L)^{1-y_L}$, $y_i =$ 0 or 1. This distribution function

belongs to (1.2.1) family, and the link function $g\left(P_i\right)=\log\frac{P_i}{1-P_i}, P_i=E\left(Y_i\right)$ is known as *logit function*. See Section 2 for a further development. If we assume

$$g\left(p\right)=\mathbf{x}^T\beta \Rightarrow p=\frac{\exp\left(\mathbf{x}^T\beta\right)}{1+\exp\left(\mathbf{x}^T\beta\right)}$$

The logit model is discussed in Section 2.

1.4 Estimation and inference for GLIM

Recall (1.2.2) and the likelihood (1.2.3). For the GLIM discussed above the log-likelihood, from (1.2.3) is reduced to

$$l(\theta;y)=\sum_{i=1}^{n} y_i b(\theta_i)+\sum C(\theta_i)+\sum d(y_i) \qquad (1.4.1)$$

and from (1.2.7)

$$\mu_l=E(y_l)=-\frac{c'(\theta_l)}{b'(\theta_l)} \qquad (1.4.2)$$

The link function is

$$g(\mu_l)=\mathbf{x}_l^T\beta=n_l \qquad (1.4.3)$$

The solution of the equation $\frac{\partial l}{\partial\theta}=0$ is equivalent to the solution of $\frac{\partial l}{\partial\beta}=0$ (see Cox and Hinkely 1974, chapter 9 for details).

Therefore, under the regularity conditions, assumed for the evaluation of MLE, the evaluation of MLE seems to be an easy story. This is not the case. The MLE can be evaluated only adopting numerical methods, the most well known being the Newton-Raphson. The theoretical insight of such a choice is beyond the target of this paper. However other choices have theoretical implementation. Following Kitsos and Edler (1997) we choose a "large" set where the solution possibly lies, adopting bisection method for 2 or 3 steps and getting an initial guess in the neighbourhood of the solution. The Newton-Raphson method can be applied and convergence is "fast" with no theoretical implementations.

For (1.4.1) the following can be proved.

Proposition 1.4.1. For the log-likelihood (1.4.1) the score function can be evaluated as

$$U_i=\frac{\partial l_i}{\partial\beta_j}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)x_{ij}}{Var(Y_i)}\left(\frac{\partial\mu_i}{\partial n_i}\right) \qquad (1.4.4)$$

Proposition 1.4.2. The Fisher's information matrix equals

$$\mathbf{I} = E(\mathbf{U}\mathbf{U}^T) = (\mathbf{I}_{jk}) = -E\left(\frac{\partial^2 l}{\partial\beta_j \partial\beta_k}\right)$$

$$\mathbf{I}_{jk} = \sum_{i=1}^{n} \frac{x_{ij}x_{ik}}{Var(Y_i)}\left(\frac{\partial\mu_i}{\partial n_i}\right)^2 \qquad (1.4.5)$$

Notice that Fisher's information is written as

$$\mathbf{I} = \mathbf{X}^T\mathbf{W}\mathbf{X} \qquad (1.4.6)$$

With $$\mathbf{W} = diag\left\{\frac{1}{Var(Y_i)}\left(\frac{\partial\mu_i}{\partial n_i}\right)^2\right\} \qquad (1.4.6a)$$

To solve the equation $$U = \frac{\partial l}{\partial\beta} = 0 \qquad (1.4.7)$$

the Newton-Ramphson iterative scheme of the following form is adopted

$$\beta_{\nu+1} = \beta_\nu - \left(\frac{\partial U}{\partial\beta}\right)^{-1}_{\beta=\beta_\nu} U_\nu \qquad (1.4.8)$$

But, considering the following steps with E being the “expectation”

$$\frac{\partial U}{\partial\beta} = \frac{\partial^2 l}{\partial\beta_l\partial\beta_k} = \frac{\partial l}{\partial\beta_l}\frac{\partial l}{\partial\beta_k} \Rightarrow E\left\{\frac{\partial l}{\partial\beta_l}\frac{\partial l}{\partial\beta_k}\right\} = \mathbf{I}_{lk} \Rightarrow -E\left[\frac{\partial^2 l}{\partial\beta_l\partial\beta_k}\right] \text{(Fisher's Information)}$$

Therefore (1.4.8) is reduced to $$\beta_{\nu+1} = \beta_\nu + \mathbf{I}_\nu^{-1}U_\nu \qquad (1.4.9)$$

The Fisher's information matrix is evaluated at $\beta=\beta_\nu$ and actually is replaced by its estimate. Multiplying both sides of (1.4.9) by I_ν we get

$$\mathbf{I}_\nu\beta_{\nu+1} = \mathbf{I}_\nu\beta_\nu + U_\nu \qquad (1.4.10)$$

Recall (1.4.6) and notation (1.4.6a) and introduce the notation

$$Z_i = \sum x_{lk}\beta_k + (y_i - \mu_i)\frac{\partial n}{\partial\mu_i} \qquad (1.4.11)$$

with β_κ at ν-iteration then (1.4.10) is reduced to

$$\mathbf{X}^T\mathbf{W}\mathbf{X}b_\nu = \mathbf{X}^T\mathbf{W}_Z \qquad (1.4.12)$$

Relation (1.4.12) is equivalent to iterative weighted least square (WLS), i.e.

Proposition 1.4.3 The MLE for GLIM is obtained by WLS.

As it is known in GLIM the MLE and OLS coincide. Therefore Proposition 1.4.3 is the equivalent of this result.

As far as statistical inference is concerned we note that the score function U has the multivariate normal N(0,I), therefore

$$\mathbf{U}^T\mathbf{I}^{-1}\mathbf{U} \sim \chi_p^2 \tag{1.4.13}$$

with **I** being Fisher's information matrix. The chi-square distribution plays a central role to the GLIM theory (see expressions (1.4.18), (1.4.19) below).

Now, if we expand the score function for the parameter β around b, by Taylor expansion then

$$U(\beta) = U(b) + H(b)(\beta - b) \tag{1.4.14}$$

with H(b) being the Hessian matrix, i.e. the matrix of second derivatives of log-likelihood at β=b. Then it can be proved that

$$\mathbf{I} = E(UU^T) = E(-H) \tag{1.4.15}$$

The following statistic (actually the "Euclidean" distance of b from β, that is how "far" is the estimate from the true parameter)

$$(b-\beta)^T I(b-\beta) \sim \chi_p^2 \tag{1.4.16}$$

is known as the *Wald statistic*. In practice is used to make statistical inference about β.

The adequacy of the model is assessed by the *likelihood ratio* statistic λ which equals to

$$\lambda = \frac{L(b_{MLE}; y)}{L(b; y)} \tag{1.4.17}$$

where b_{MLE} the MLE of β, $L(\cdot)$ the likelihood function. Therefore λ offers a measure of *goodness of fit* as the ratio of the maximal model and the model of interest. Moreover

$$\log\lambda = \ell(b_{MLE}; y) - \ell(b; y) \tag{1.4.17a}$$

Nedler and Wedderburn (1972) defined the *deviance D* as

$$D = 2\log\lambda = 2\left[\ell(b_{MLE}; y) - \ell(b; y)\right] \sim \chi^2_{n-p} \tag{1.4.18}$$

As theoretically the chi-square is a measure of distance by (1.4.18) we measure the distance of the log-likelihood.

If the model is poor D will be larger predicted by χ^2_{n-p}. Theoretical D follows a non-central χ^2, but this is beyond this paper. Thus if a model –with p parameters– provides a good description of the collected n observation so that $D \sim \chi^2_{n-p}$, we expect

$$D \sim N - P \tag{1.4.18a}$$

The deviance D can be a helpful tool for a hypothesis testing. Indeed if we are interesting in testing $H_0 : \beta = \beta_0$ vs. $H_1 : \beta \neq \beta_1$

with $\beta_0 = \left(\beta_1,...,\beta_q\right)^T, \beta_1 = \left(\beta_1,...,p_p\right)^T$, 2<p<n

Proposition 1.4.4. We can test H_0 versus H_1 by adopting the difference of deviances under H_0 and $H_1, say D_0, D_1$ respectively.

Indeed:

$$\begin{aligned} D = D_0 - D_1 &= 2\left[\ell(b_{MLE}; y) - \ell(b_0; y)\right] - 2\left[\ell(b_{MLE}; y) - \ell(b_1; y)\right] \\ &= 2\left[\ell(b_1; y) - \ell(b_0; y)\right] \sim \chi^2_{p-q} \end{aligned} \tag{1.4.19}$$

Thus if D has been evaluated greater than the upper tail 100% point of the χ^2_{p-q} we reject H_0 in favour of H_1 – i.e. even though β_1 has too many parameters does not fit the collected data set satisfactory. Notice that F distribution can be used only for model involving normal distribution. In such a case

$$F = \frac{D_0 - D_1}{p-q} \Big/ \frac{D_1}{n-p} \sim F_{p-q,n-p} \tag{1.4.20}$$

If H_0 is not correct the calculated value of F will be larger than the value of the corresponding $F_{p-q,n-p}$ distribution, at the predefined level.

2. Logistic Regression

The logistic regression is a particular case of a GLIM. We are going to discuss the simple logistic model next.

2.1 Introduction

There are real life problems were interest is focused on the number of "successes" or "failures" within a (sub) group of observations. The idea was extended to various other problems, number of employment/unemployment, to prefer or not a product (or procedure or course or candidate etc). This is the typical situation were for years the ordinary least square method (OLS) was adopted, since Gauss in problems of geodesy and latter, since Kendall adopted it to social sciences, economy etc.

Let us formulate the problem. We are working with generalized linear models (GLIM), were the response is a binary variable declaring "success" or "failure" as was above discuss. Let W be such a variable:

$$W = \begin{cases} 1 & \text{if the outcome is success} \\ 0 & \text{if the outcome is failure} \end{cases} \qquad (2.1.1)$$

With $$P(W=1) = \pi = 1 - P(W=0) \qquad (2.1.2)$$

Certainly any number could declare success or failure say -1 and 1, 5 and 15 BUT we are using 1 and 0 (and the statistical packagers only these values realize) as we are NOT interested in W eventually, but on the summation of values W.

Actually if they are observed in such independent random variables with assessing probability then the joint probability is evaluated as.

$$\prod_{i=1}^{n} \pi_i^{w_i} (1-\pi)^{1-w_i} = \exp\left\{ \sum_{i=1}^{n} w_i \log\left(\frac{\pi_i}{1-\pi_i} \right) + \sum_{i=1}^{n} \log(1-\pi_i) \right\} \qquad (2.1.3)$$

Notice that (2.1.3) belongs to the exponential family of models discussed in section (1.2) assuming that π_i's are equal, i.e.

$$W = \begin{cases} 1, \text{ success with probability } \pi \\ 0, \text{ failure with probability } 1-\pi \end{cases} \qquad (2.1.4)$$

In such a case if we define the random variable as

$$Y = \sum_{i=1}^{n} w_i = \text{number of successes} \qquad (2.1.5)$$

it is a binomial distribution with probability density function defined as

$$P(Y=y) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \quad y = 0, 1, \ldots, n \qquad (2.1.6)$$

So we can classify the data as in Table 2 if N independent random variables $Y_1, Y_2, \ldots Y_N$ corresponding to the number of successes/failures are assumed.

Table 2. Frequencies of N binomial distributions

	Subgroup			
	1	**2**	**…**	**N**
Successes	Y_1	Y_2	…	Y_N
Failures	$n_1 - Y_1$	$n_2 - Y_2$	…	$n_N - Y_N$
Totals	n_1	n_2	…	n_N

The most typical case in practice is to consider two subgroups. Examples are employment/unemployment for men/women, getting or not a characteristic for patient /not patient, etc.

2.2 Generalized Linear Models

The target is to "describe", to "analyze" the proportion of successes $P_i = \frac{Y_i}{n_i}$, in each subgroup (see Table 2.1) in terms of e.g. factor levels, sex age etc acting as explanatory variables X_i which "describe" the subgroup. That is we are interested in assuming that there is a link function which is modelling the probabilities π_i as

$$g(\pi_i) = \mathbf{X}_i^T \beta \qquad (2.2.1)$$

with β being a vector of parameters.

Trivially if $g(\pi_i) = \mathbf{X}_i^T \beta$ then $\pi = g^{-1}(\mathbf{X}^T \beta)$ and as π needs to be within [0,1] we assume the existence of a cumulative distribution function (cdf) F such that

$$\pi = g^{-1}(\mathbf{x}^T \beta) = F(t) = \int_{-\infty}^{t} f(w) dw \qquad (2.2.2)$$

as by the definition of the cdf a distribution function f exists such that (2.2.2) to be true. This probability density function f(w) is called tolerance distribution. As there are various tolerance distributions we shall focus our attention on

$$f(w) = \frac{\beta_1 \exp(\beta_0 + \beta_1 w)}{[1 + \exp(\beta_0 + \beta_1 w)]^2} \qquad (2.2.3)$$

known as *logit* or *logistic model*.

From (2.2.3) and (2.2.2) we have

$$\pi = \int_{-\infty}^{x} f(w) dw = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \qquad (2.2.4)$$

Therefore the link function (see (2.2.1)) is

$$g(\pi) = \log \frac{\pi}{1 - \pi} = \beta_0 + \beta_1 x \qquad (2.2.4\alpha)$$

This link function, $\log \frac{\pi}{1 - \pi}$,is known as *logit function* or *logit transformation*.

2.3 MLE for the Logit

We are fitting the logistic model of the form

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_1)}{1 + \exp(\beta_0 + \beta_1 x_1)}$$

So $$\log \frac{\pi_i}{1-\pi_i} = \beta_0 + \beta_1 x_i \Rightarrow \log\left(1-\pi_i\right) = -\log\left[1+\exp\left(\beta_0 + \beta_1 x_i\right)\right] \qquad (2.3.1)$$

From (2.1.3) and (2.3.1) we get the log-likelihood function equal to

$$\ell = \sum_{i=1}^{n}\left\{ y_i\left(\beta_0 + \beta_1 x_i\right) - n_i \log\left[1+\exp\left(\beta_0 + \beta_1 x_i\right)\right] + \log\binom{n_i}{y_i}\right\} \qquad (2.3.2)$$

Therefore the score functions with respect to β_0 and β_1 are evaluated as

$$\begin{aligned} U_0 &= \frac{\partial \ell}{\partial \beta_0} = \sum_{i=1}^{n}\left(y_i - n_i\pi_i\right) \\ U_1 &= \frac{\partial \ell}{\partial \beta_1} = \sum_{i=1}^{n} x_i\left(y_i - n_i\pi_i\right) \end{aligned} \qquad (2.3.3)$$

Therefore Fisher's 2x2 information matrix $\mathbf{I}$ is evaluated equal to

$$\mathbf{I} = \begin{pmatrix} \sum n_i\pi_i\left(1-\pi_i\right) & \sum n_i x_i\left(1-\pi_i\right) \\ \sum n_i x_i\pi_i\left(1-\pi_i\right) & \sum n_i x_i^2\pi_i\left(1-\pi_i\right) \end{pmatrix} \qquad (2.3.4)$$

Recall that MLE are obtained through the iterative scheme

$$\mathbf{I}_{\nu-1}\beta_\nu = \mathbf{I}_{\nu-1}\beta_{\nu-1} + U_{\nu-1} \qquad (2.3.5)$$

see (1.4.10). Therefore an initial guess should be provided. To "feed" (2.3.5) initial guesses might be used - take $\beta_{0,0} = 0,\ \beta_{1,0} = 0$. The estimated variance-covariance matrix is $\left[\mathbf{I}\left(\beta_\nu\right)\right]^{-1}$, the fitted values then are $\hat{y}_i = n_i\hat{\pi}_i$ and

- the log-likelihood ratio statistic is

$$\log\lambda = \sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{\hat{y}_i}\right) + \left(n_i - y_i\right)\log\left(\frac{n-y_i}{n-\hat{y}_i}\right)\right] \qquad (2.3.6)$$

As $D = 2\log\lambda$ see (1.4.18) and $\hat{y}_i = n_i\hat{\pi}_i$ we get from (2.3.6) that the deviance D is

$$D = 2\sum_{i=1}^{n}\left[y_i \log\left(\frac{y_i}{n\hat{\pi}_i}\right) + \left(n_i - y_i\right)\log\left(\frac{n-y_i}{n-n_i\hat{\pi}_i}\right)\right] \qquad (2.3.7)$$

Notice that (2.3.6a) does not involve any nuisance parameters and recall that $D \square \chi^2_{n-p}$, see (1.4.18a).

2.4 Goodness of fit

Recall that Pearson's χ^2 statistic is defined as

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{2.4.1}$$

That is (2.4.1) offers a measure of distance between the observed (O) and the expected values (E). For the cells of Table 2.1 this is clear.

For the model under investigation

$E(Y_i) = n_i\pi_i$ and $Var(Y_i) = n_i\pi_i(1-\pi_i)$

- so the weighed sum of squares

$$WSS = \sum_{i=1}^{n} \frac{(y_i - n\pi_i)^2}{n_i\pi_i(1-\pi_i)^2} \tag{2.4.2}$$

has a meaning to attract interest.

Proposition 2.4.1 The WSS is equivalent to χ^2. Indeed, if we consider (2.4.1):

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{n} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i} + \sum_{i=1}^{n} \frac{\left[(n_i - y_i) - n_i(1-\pi_i)\right]^2}{n_i(1-\pi_i)} \\ &= \sum_{i=1}^{n} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1-\pi_i)}(1-\pi_i+\pi_i) = WSS \end{aligned} \tag{2.4.3}$$

Therefore when is calculated at the estimated values is

$$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)} \tag{2.4.4}$$

A very nice result for the χ^2, measure of distance between observed and expected values is that

Proposition 2.4.2 Relation (2.3.6) is equivalent to (2.4.4). Therefore the evaluated deviance for the logit model coincided with the calculated χ^2_{cal} values.

Indeed:

We shall apply the Taylor series expansion of $K \log \frac{K}{L}$ about K=L, of the form

$$K \log \frac{K}{L} = (K-L) + \frac{1}{2}\frac{(K-L)^2}{L} + \ldots$$

We apply this Taylor expansion for the two terms of D with $K = Y_i, L = n\hat{\pi}_i$ and $K = (n_i - y_i)$ and $L = n_i - n_i\hat{\pi}_i$ as follows

$$\begin{aligned} D &= 2\sum_{i=1}^{n}\left[y_i \log \Big/ \left(\frac{y_i}{n_i\hat{\pi}_i} \right) + (n_i - y_i) \log \ell \left(\frac{n_i + y_i}{n_i - n_i\hat{\pi}_i} \right) \right] \\ &\cong 2\sum_{i=1}^{n}\left[(y_i - n_i\hat{\pi}_i) + \frac{1}{2}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i} + \ldots \right] + \\ &+2\sum_{i=1}^{n}\left[\left\{ \left[(n_i - y_i) - (n_i - n_i\hat{\pi}_i) \right] + \frac{1}{2}\frac{\left[(n_i - y_i) - (n_i - n_i\hat{\pi}_i)^2 \right]^2}{n_i - n_i\hat{\pi}_i} + \ldots \right\} \right] \\ &\cong \sum_{i=1}^{n}\frac{(y_i - n_i\hat{\pi}_i)^2}{n_i\hat{\pi}_i(1-\hat{\pi}_i)} = \chi^2_{cal} \end{aligned} \qquad (2.4.5)$$

2.5 Extensions

Let us now discuss another very important index in Logistic analysis, the *odds ratio*. First we define the distributional properties of the dependent variable[1], which is a dichotomous variable Y taking the value of 1 with probability Θ and the value of 0 with probability 1-Θ. Such a random variable has a simple discrete probability distribution given as

$$\Pr(Y_i, \Theta_i) = \Theta_i^{Y_i}(1-\Theta_i)^{1-Y_i} \qquad (2.5.1)$$

Given the mutually independent $Y_1, Y_2, \ldots, Y_n$, the likelihood function of (2.5.1) is the product of the marginal distributions for the Y_i 's. Specifically

$$L(Y;\Theta) = \prod_{i=1}^{n} \Pr(Y_i;\Theta_i) = \prod_{i=1}^{n}\left(\Theta_i^{Y_i}(1-\Theta_i)^{1-Y_i} \right) \qquad (2.5.2)$$

where $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)$.

In our sample the first n_1 out of n observations have the characteristic under investigation (employment – unemployment, having information – not having

[1] For more details on the properties and applications of logistic regression see Halkos (2006), Kleinbaum (1994), Hosmer and Lemeshow (1989), Kleinbaum *et al.* (1999), Hair et al. (1998), Sharma (1996).

information etc) and so $Y_1=Y_2=\ldots=Y_{n_1}=1$ while the rest of the observations do not and so $Y_{n_1+1}=Y_{n_1+2}=\ldots=Y_n=0$. This means that expression (2.5.2) becomes

$$L(Y;\Theta)=\left(\prod_{i=1}^{n_1}\Theta_i\right)\left[\prod_{i=n_1+1}^{n}(1-\Theta_i)\right] \qquad (2.5.3)$$

If $X_i=(X_{i1}, X_{i2}, \ldots, X_{ik})$ the set of values of the k independent variables $X_1, X_2, \ldots, X_k$ specific to individual i then the logistic model assumes that between Θ_i and X_{ij}'s a specific form exists which is given by

$$\Theta_i=\frac{1}{1+\exp\left[-\left(\beta_0+\sum_{j=1}^{k}\beta_j X_{ij}\right)\right]} \qquad i=1,2,\ldots,n \qquad (2.5.4)$$

Obviously β_j are unknown coefficients to be estimated by regression. Replacing Θ_i in (3) we derive the likelihood function as

$$L(Y;\beta)=\frac{\prod_{i=1}^{n_1}\exp^{(\beta_0+\sum_{j=1}^{k}X_{ij})}}{\prod_{i=1}^{n}\left[1+\exp^{\left(\beta_0+\sum_{j=1}^{k}\beta_j X_{ij}\right)}\right]} \qquad (2.5.5)$$

Although we assume an unconditional maximum likelihood function that could lead to biased estimates of β's as our data size is large this potential problem is not so serious.

The regression coefficients β's of the proposed logistic model quantifies the relationship of the independent variables to the dependent variable involving the parameter called the *Odds Ratio* (OR).

Definition 2.5.1. As odds we define the ratio of the probability that implementation will take place divided by the probability that implementation will not take place.

That is $$\text{Odds}\,(E\,|\,X_1, X_2, \ldots, X_n)=\frac{\Pr(E)}{1-\Pr(E)} \qquad (2.5.6)$$

Instead of minimizing the squared deviations as in a multiple regression, logistic regression maximizes the likelihood that an event will take place.

$$\ln\frac{\Pr(E)}{1-\Pr(E)}=\beta_0+\beta_1X_1+\beta_2X_2+\ldots+\beta_kX_k \qquad (2.5.7)$$

or $$\Pr(E)=\frac{1}{1+\exp^{-(\beta_0+\sum_{i=1}^{k}\beta_i X_{ij})}} \qquad (2.5.8)$$

where P is the probability of having the characteristic under investigation given the independent variables X_1, X_2,…, X_k. Equation (2.5.8) models the log of the odds as a linear function of the independent variables and it is equivalent to a multiple regression equation with log of the odds as the dependent variable.

The logit form of the model is a transformation of the probability Pr(Y=1) that is defined as the natural log odds of the event E(Y=1). That is

$$\text{logit } [\Pr(Y=1)]=\ln[\text{odds } (Y=1)]=\ln\left[\frac{\Pr(Y=1)}{1-\Pr(Y=1)}\right] \qquad (2.5.9)$$

Let us consider the general case, where the dichotomous response variable Y, denotes whether (Y=1) or not (Y=0) the characteristic under investigation (employment – unemployment, having information – not having information etc) is linked with the k regression variables $X=(X_1, X_2, \ldots, X_k)$ via the logit equation

$$P(Y=1)=\frac{\exp\left\{\beta_0+\sum_{k=1}^{K}\beta_k X_k\right\}}{1+\exp\left\{\beta_0+\sum_{k=1}^{K}\beta_k X_k\right\}} \qquad (2.5.10)$$

This is equivalent to

$$\text{logit } \Pr(Y=1|\, X)= \beta_0+\sum_{k=1}^{K}\beta_k X_k$$

With this formulation we have the benefit that the *relative risk* (RR) for individuals having two different sets X′ and X of risk variables is

$$RR=\frac{P(X')\left[1-P(X)\right]}{P(X)[1-P(X')]}=\exp\left\{\sum_{i=1}^{K}\beta_i(X_i'-X_i)\right\} \qquad (2.5.11)$$

It is essential that the RR of the k regressors (RR_k) influences the RR of the k+1 regressors, RR_{k+1} as in relation (2.512) below, due to the following Theorem 2.5.1.

Theorem 2.5.1

Let the relative risk of as above in (2.5.11) is

$$RR_K=\sum_{i=1}^{K}\beta_i\delta_i \qquad \text{with} \qquad \delta_i=X_i'-X_i$$

Then if a variable is added the relative risk of the k+1 variable equals the k-variables relative risk times the new variable relative risk, ie.

$$RR_{k+1} = RR_k\, r_{k+1} \qquad (2.5.12)$$

Proof

Indeed

$$RR_{K+1} = \exp\left\{\sum_{i=1}^{K+1}\beta_i\delta_i\right\} = \exp\left\{\sum_{i=1}^{K}\beta_i\delta_i + \beta_{K+1}\delta_{K+1}\right\} = \exp\left\{\sum_{i=1}^{K}\beta_i\delta_i\right\}\exp\{\beta_{K+1}\delta_{K+1}\} = RR_K \cdot r_{K+1}$$

Where the definition of r_{K+1} is obvious. So the relative risk of the incoming K+1 variable is

$$r_{K+1} = \frac{RR_{K+1}}{RR_K} \quad .$$

Now suppose that v is an indicator variable denoting whether or not (v=1 or v=0) someone is sampled. If we denote by

$$\mu_1 = P(v=1 \mid Y=1)$$
$$\mu_0 = P(v=1 \mid Y=0) \qquad (2.5.13)$$

the probability that the one with the characteristic is included in the study (μ_0) and the probability that the one without the characteristic participates as a control then the conditional relative risk influences only the constant term β_0, while the rest part of RR_K remains "invariant", due to the following Theorem 2.5.2.

Theorem 2.5.2

The relative risk defined as the conditional probability that a person has the characteristic, given that has risk variables X and also was sampled from the case-control studies is of the form (2.5.10) with constant term β_0' equal to:

$$\beta_0' = \beta_0 + \log\frac{\mu_1}{\mu_0} \qquad (2.5.14)$$

Proof:

From Bayes' theorem

$$P(Y=1 \mid v=1,X) = \frac{P(v=1/Y=1,X)P(v=1/X)}{P(v=1/Y=0,X)P(y=0/X) + P(v1/Y=1,X)P(Y=1/X)} =$$

$$= \frac{\mu_1\exp\left\{\beta_0 + \sum_{i=1}^{k}\beta_i X_i\right\}}{\mu_0 + \mu_1\exp\left\{\beta_0 + \sum_{i=1}^{k}\beta_i X_i\right\}} = \frac{\frac{\mu_1}{\mu_0}\exp\left\{\beta_0 + \sum_{i=1}^{k}\beta_i X_i\right\}}{1 + \frac{\mu_1}{\mu_0}\exp\left\{\beta_0 + \sum_{i=1}^{k}\beta_i X_i\right\}} = \frac{\exp\left\{\beta_0' + \sum_{i=1}^{k}\beta_i X_i\right\}}{1 + \exp\left\{\beta_0' + \sum_{i=1}^{k}\beta_i X_i\right\}}$$

With β_0' as in (2.5.14).

Let us now estimate the marginal influence of extra regressors. It is known that the percentile point L_p, for the cumulative distribution function F(x) is defined as $F(L_p)=p$. Therefore for the logistic $\Lambda(\cdot)$, say, we have

$$\Lambda\left(L_p : \theta\right) = \left\{1 + \exp\left[-(\theta_0 + \theta_1 L_p)\right]^{-1}\right\} = p \qquad (2.5.15)$$

Solving equation (2.5.15) for L_p it can be evaluated that

$$L_p = -\theta_1^{-1}\left[\theta_0 + \ln\left(p^{-1} - 1\right)\right] \qquad (2.5.16)$$

From (2.5.16) with p=0.5, i.e. for the median $M=L_{0.5}$ it is obtained equal to

$$M = -\frac{\theta_0}{\theta_1} \qquad (2.5.17)$$

In economical analysis the median M, although a biased estimator, is sometimes, more useful than the unbiased mean, as it offers a better measure of the location of the data. Median, even under the minimax criterion, is not selected compared to the mean. But as it remains invariant if the "end-points" change drastically it is a better location measure in many cases of the economical data analysis. Consider that the value of L_p in (2.5.16) is a function of the parameter, i.e. $L_p=L_p(\theta)$, with $\theta=(\theta_0, \theta_1)$. With the above discussion in mind, we state and prove the following proposition.

Proposition 2.5.1:
The minimization of the variance of the percentile point L_p, $Var(L_p)$, is (approximately) equivalent to minimize the

$$C^T I_F^{-1}(\theta) C \qquad (2.5.18)$$

With C=(1,Lp) and $I_{F^*}^{-1}$ the (*average per observation*) Fisher's information and $\theta=(\theta_0, \theta_1)$.

Proof:
It is known, in principle for the exponential family of models, with response η it is

$$I_F(\theta) = \sigma^{-2}(\nabla\eta)(\nabla\eta)^T = vv^T \qquad (2.5.19)$$

where the definition of the vector v is obvious, and v^T is the transpose of v, and ∇ is the "grand" vector asn usually. Moreover in many of the nonlinear problems the covariate u and the parameter θ appears "together" linearly, i.e. $\eta = \eta(\theta^T u)$. In such a case

$$\nabla\eta = \left[w\left(\theta^T u\right)\right]^{1/2} u \qquad \text{with} \qquad w(z) = \left[\frac{\partial\eta}{\partial z}\right]^2, \quad z = \theta^T u .$$

That is Fisher's information matrix is

$$I_F(\theta) = \sigma^{-2} w(\theta^T u) u u^T \qquad (2.5.20)$$

With σ^2 the involved variance, practically a function of the unknown parameters we are asked to estimate. Obviously the Logistic model obeys on the "*intrinsic*" linearity described above. Then for the MLE of θ, $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1)$ and for n_i the number of observation at u_i i=1, 2 and $\Sigma n_i=n$ large, then from (2.5.20), we have

$$Var(\hat{\theta}) = Var(\hat{\theta}_0, \hat{\theta}_1) = \left\{\sum \Lambda(u_i)\left[1-\Lambda(u_i)\right]\mu_i u_i^T n_i\right\}^{-1} \qquad (2.5.21)$$

Based on the fact that $I_F^{-1}(\theta) = Var(\theta)$ and u_2 =(1, u_2), i=1,2. From (2.5.21) the evaluation of $Var(\hat{L}_p)$ equals

$$Var(L_p) = (\nabla L_p)^T Var(\hat{\theta})(\nabla L_p) = C^T I_{F^*}^{-1}(\theta) C \qquad (2.5.22)$$

Where $I_F^{-1}(\theta) = n^{-1}\sum I(\theta)$, with I(θ) Fisher's information and $I_{F^*}^{-1}(\theta)$ the average per observation information matrix.

2.6 Discussion on the Odds Ratio

Consider the typical scheme in logit problems

Risk Group	Case	Control
Characteristic XX	P_1	P_0
Non – XX	$1-P_1$	$1-P_0$

The probability P_1 represents the probability that a "sample case member" (e.g. patient exposed to the risk factor, unemployed attended certain seminar, company adapted an innovation process, etc) exposed to the risk factor. The probability P_0 is the probability that a simple "control member" is exposed to the risk factor.

The *odds ratio* (OR) *or relative risk* (RR) is defined as

$$\text{RR or } OR = \frac{P_1/(1-P_1)}{P_0/(1-P_0)} = \frac{P_1(1-P_0)}{P_0(1-P_1)} \qquad (2.6.1)$$

A χ^2 test of the null hypothesis H_0: OR=1 vs H_1: OR≠1 is identical to a test of equality of the two proportions, i.e.

$$H_0: P_0 = P_1 \text{ vs } H_1: P_0 \neq P_1 \qquad (2.6.2)$$

The crude odds ratio can be biased due to population heterogeneity caused by confounding factors, associated with the response. Significance tests and confidence intervals for the crude OR were introduced by Mantel and Haenszel (1959) in their pioneer paper.

In the typical case where

$$P(case/x) = \frac{\beta_0 + \beta_1 x}{1 + \exp(\beta_0 + \beta_1 x)} \qquad (2.6.3)$$

With the dichotomous variable X=1 for the high risk (characteristic – XX) and X=0 for the low-Risk (non - XX), the odds ratio is given by

$$OR = \exp(\beta_1) = RR \cong \exp(\hat{\beta}_1)$$

underlying that the above notation was adopted.

- That is the Relative Risk (RR) or (OR) testing hypothesis H_0: RR=1 is equivalent to H_0: $\beta_1=0$.

When interest is focused on the trichotomous factor XX versus XY versus YY we introduce two dummy variables X_1, X_2 defined as

$X_1=1$ if XX holds
$=0$ otherwise

$X_2=1$ if XY holds
$=0$ otherwise

Therefore YY is the reference class. The equation of the model is

$$\text{Logit } P(\text{case} \mid x_1, x_2; z)=\beta_0+\beta_1 X_1 +\beta_2 X_2 +\gamma Z$$

In such a case there are two different relative risks or odds ratios of the form

$$OR_1=RR_1=\exp(\beta_1) \approx \exp(\hat{\beta}_1)$$
$$OR_2=RR_2=\exp(\beta_2) \approx \exp(\hat{\beta}_2) \qquad (2.6.4)$$

The OR_1 provides a measure of the relative risk or odds ratio of XX versus YY, while OR_2 is a measure of OR or RR of XY versus YY, adjusted for the confounders of Z. When the variable Z is included in the logit model with one variable still the $OR=\exp(\beta_1)$.

3. The Greek Innovation activities

In the following the Greek innovation activities are discussed through the logit model and an empirical application, due a survey we performed to the Greek firms. We explain this analysis below.

3.1 DATA

The data used in this study rely on our survey to the Greek enterprises through 'ARCHIMEDES' program, mentioned above. The target population of the survey is the total population of 63.000 firms, included in the database of ICAP SA (the largest Business Information and Consulting firm in Greece). The constructed questionnaire is based on Community Innovation Survey (CIS) while the methodological basis of this survey is provided by the "Oslo manual", a joint publication of Eurostat and OECD. The selection of the sample is based on proportional stratified random sampling. The analysis focuses mainly on the small and medium-sized manufacturing enterprises and on services enterprises employing 20 or more employees. To avoid any bias resulting from response behaviour, a non-response analysis followed up.

The main statistical unit for the survey is the enterprise. In general, innovation activities and decisions usually take place at the enterprise level, which leads to the enterprise being used as the statistical unit. The following industries are included in the target population of the survey:

- manufacturing (NACE 15-37),
- electricity, gas and water supply (NACE 40-41),
- wholesale trade (NACE 51),
- transport, storage and communication (NACE 60-64),
- financial intermediation (NACE 65-67),
- computer and related activities (NACE 72),

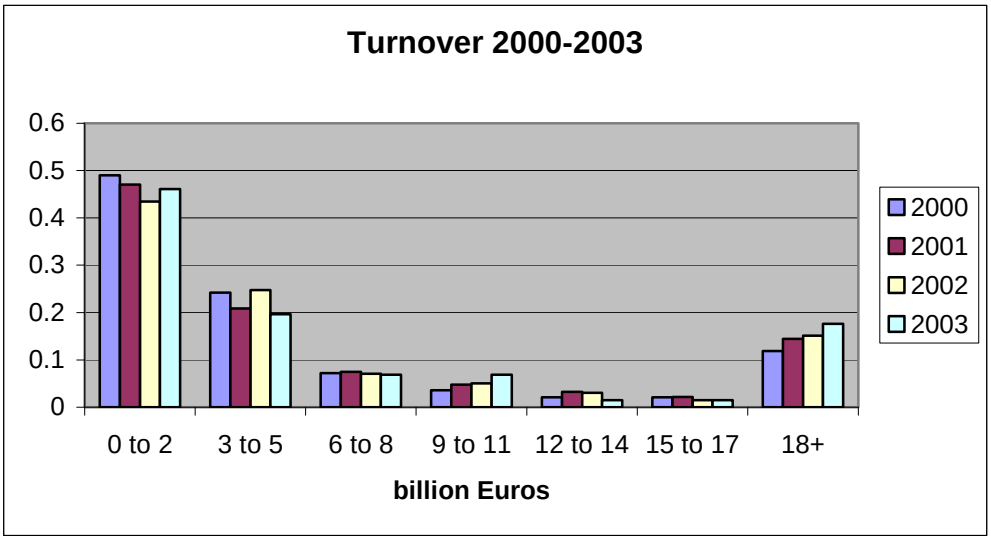
- architectural and engineering activities (NACE 74.2),
- technical testing and analysis (NACE 74.3),
- research and development (NACE 73),
- construction (NACE 45),
- motor trade (NACE 50),
- retail trade (NACE 52),
- legal, accounting, market research, consultancy and management services (NACE 74.1),
- advertising (NACE 74.4),
- labour recruitment and provision of personnel (NACE 74.5),
- investigation hotels and restaurants (NACE 55),
- renting of machinery and equipment without an operator (NACE 71).

The time period covered by the survey was 2000-2003 inclusive. The reference period of the survey is the year 2003. Following the European classification, the size-classes used are 20-49 employees, 50-249 employees, 250 + employees.

3.2. Empirical findings

In collecting primary data, we have used a questionnaire to several sectors in Greek enterprises. The data-set was collected from 279 Greek enterprises from various sectors and several areas-prefectures of the country. From the data collected it can be seen that the R&D activities are mainly related to big firms. Also Figure 1 presents the turnover of the sampled firms and for the years 2000-2003. In the vertical axis we have the percentages while in the horizontal axis we have the turnovers in billion €.

Figure 1: Turnover of sampled firms for the years 2000-2003



The share of innovative enterprises with 10 or more employees in the service industry increased significantly from 11,1% in 2001 to 15,5% in 2002 and the 31,9% in 2003. The analysis of data from Greece and other European countries reveals that the development of innovations does not require the development of R&D activities within the enterprise. In Greece, 59.7% in 2001, 62.3% in 2002 and 64,7% in 2003 of the innovative manufacturing enterprises carry out R&D. In the service sector, the percentage is much lower. Specifically, 8.3% in 2001, 13% in 2002 and 16.1% in 2003 of the innovative enterprises with 10 or more employees carry out R&D.

They are directed towards to obtain the machinery and equipments accounting around to 58%. Education is the main source for innovation activities in Greek enterprises accounting approximately to 48% while the introduction of innovation in the market accounts to approximately 27% and the planning for production-distribution and other activities to 25 % respectively.

The following findings are based on the analysis of the collected 279 questionnaires of our sample survey (Kitsos *et al.*, 2006). We are now at the final stage of checking our full data set. We are proceeding our calculations on the double checked 279 responses, and we shall extent this discussion to the final data set. Our empirical results concern the total number of women employees by age intervals (namely 20-30, 31-40, 41-50, 51-60 and more than 60), women in product innovation, women in process innovation, position in firm and equality in job enrichment, in salary, in education–training and in promotion. Specifically,

- A small increase in the percentage of women in the workplace has occurred from 32% in 2000 to almost 37,5% in 2003.
- A small increase in the percentage of women graduates in the workplace has occurred from almost 33% in 2000 to almost 37% in 2003.
- Women employees between 20-30 ages (compared to 31-50 ages) are the majority of the women workforce and a small increase in the percentage has occurred from 59% in 2000 to 61% in 2003.
- The percentage of women employees is decreased as long as the age is increased to 62% for 20-30, 23% for 31-40, 10% for 41-50, 3% for 51-60 and 2% for over 60 in 2003.
- The percentage of women as Top Managers remains almost unchanged, from 21% in 2000 to 20% in 2003 while the percentage of women as Managers is increased from 33% in 2000 to 38% in 2003.
- The percentage of women employees in product innovation development is bigger (40%) than the percentage in process innovation development (33%) and the percentage of women as Top Management in product innovation development is bigger (20 %) than the percentage in process innovation development (16%).
- The percentage of women as Managers in product innovation development is far smaller (29%) than the percentage in process innovation development (45%).
- The percentage of the firms with innovation activities, concerning all the four variables of gender equality in the workplace (job enrichment, salary, training and promotion), appears to be smaller than the percentage in the total of the firms of the sample. For example, job enrichment appears to be around 37% in product innovation and 26% in process innovation comparing to 71% of the total of firms (with and without innovation activities).

The main implications for both small and medium firms for the products innovations is related to increase the range of commodities and less to improve the quality of commodities. Regarding the finance of innovations, a small part of Greek firms has been financed either from local-regional authorities or from the central government and the European Union. Specifically, around 18% of the Greek sampled firms have been financed for innovations from the EU while the other sources of finance, like the central government and the local-regional authorities correspond to very low levels of 6% and 3% respectively.

According to the collected data, the main obstacles in the development of innovations are related to the lack of appropriate financial sources for innovation activities which accounts to 70% for small firms and only 10% for big firms, while the high risk activities account to about 74% for small firms and only 5% for big firms respectively. Another serious obstacle is related to the lack of information of new technologies and for the markets accounting 75% for small firms and 25% for medium firms. At the same time the lack of specialized staff accounts to 60 % for small firms and 40% for medium firms and the managerial inflexibilities account to 40 % for the small firms and 40% for the medium firms respectively.

3.3 Logit formulations and associated empirical results

As our main interest is in terms of the main effects we have ignored interactions. Only 3 and 4 out of the 12 explanatory variables were found to be statistically significant in influencing the implementation of product and process innovations respectively. Working with the most statistical significant variables we derive the logit form of the fitted model, which may be represented as

$$\text{logit } [\Pr(Y=1)]=\beta_0+\beta_1 \text{Turnover}+\beta_2 \text{ PAL} + \beta_3 \text{ PFW} + \beta_4 \text{ Exp} + \varepsilon_t$$

where Y denotes the dependent variable as 1 for innovations and 0 for no innovations, the beta terms are the unknown linear coefficients needed to be estimated, and ε_t is the error term, assumed from the normal distribution with mean 0 and variance 1.

Specifically the dependent variable is the answer to the question of the influence of innovations to the products with answers ranging from high influence to no influence. The high and average influences were coded as 1 and the low and no influence as 0. The explanatory variables are Turnover taking the value of 1 for a higher than 10% increase in the turnover for the period 2000-2003 and 0 in any other case, the PAL that is the *Product Average Life* which is the average life of the most important product of the firm before it is substituted or modified. It takes the values of 1 in case of less than a year, 2 in case of 1-3 years, 3 for 4-6 years, 4 for 7-9 years, 5 in case of more than 9 years and 0 if no answer). The last significant explanatory variables are the PFW that is the percentage of female workers and Exports. The results of the fitted models are presented in Table 3.

Based on the fitted model and the information provided, it can be seen that the estimated odds ratio equals to 3,219 and 1.897 for implementing product and process innovations respectively for firms which have a higher than 10% increase in turnover, with no control for the other explanatory variables. This adjusted odds ratios of 3.219 and 1.897 mean that the odds of implementing product and process innovations is about 3.219 and 1.897 times higher respectively for a firm which has an increase of more than 10% of its turnover than for a firm which has not. The Wald statistic is statistically significant, which indicates that there is statistical evidence in these data that for firms with turnover higher than 10% the probability of implementing innovation increases.

We may compute the difference $e^{\hat{\beta}_i} - 1$ which estimates the percentage change (increase or decrease) in the odds $\pi = \dfrac{\Pr(Y=1)}{\Pr(Y=0)}$ for every 1 unit in X_i holding all the other X's fixed. The coefficient of average life of the product is $\hat{\beta}_2$=0.413, which implies that the Relative Risk of this particular variable is $e^{\hat{\beta}_2}$=1.511 and the corresponding percentage change is $e^{\hat{\beta}_1}$ -1=0.511. This means that in relation to the

average life of product the odds of implementing innovations increase by 51.1% ceteris paribus. In the case of process innovation the result is 1.271 implying a 27.1% increase. In relation to turnover the odds of implanting innovations increase by 221.9% and 89.7% for product and process innovations respectively keeping constant all the rest. Similarly, the coefficient of percentage of female workers is $\hat{\beta}_3$ =-0.027 and -0.014 for product and process innovations respectively, which implies that the Relative Risk of this particular variable is $e^{\hat{\beta}_3}$=0.9733 and 0.986 and the corresponding percentage change is $e^{\hat{\beta}_2}$-1= -0.0267 and -0.014 respectively. This means that in relation to the percentage female workers the odds of implementing innovations decreases by almost 0.03% and 0.015 all other remaining fixed.

Table 3: Logistic Regression results

Dependent: Product Innovations				**Dependent: Process Innovations**	
Variables	**Estimates**	**Odds Ratio**		**Estimates**	**Odds Ratio**
Constant	0.483 (0.734) [0.391]	1.621			
Turnover	1.169 (6.071) [0.014]	3.219		0.64 (2.895) [0.089]	1.897
PAL	0.413 (8.700) [0.003]	1.511		0.240 (4.403) [0.036]	1.271
PFW	-0.027 (6.510) [0.011]	0.973		-0.014 (3.028) [0.082]	0.986
Exports				1.139 (5.252) [0.022]	3.123
Hosmer Lemeshow	8.145 [0.37]			11.135 [0.47]	
Likelihood Ratio	18.121 [0.000]			28.016 [0.000]	

Wald statistics in parentheses and P-values in brackets.

The negative sign in the coefficient for the percentage of female workers variable requires some further thoughts. It could be explained as a higher percentage of female workers corresponding to a lower implementation of innovations. This can be considered in relation to other Departments in the firm like R&D as well as to the markets that the firm operates.

The individual statistical significance of the β estimates is presented in the column Wald (Chi-square). The significance levels of the individual statistical tests (i.e. the P-values) are presented in the column P-value and correspond to Pr>Chi-square. Note that the variable average value of product is significant in all the usual statistical levels (0.01, 0.05 or 0.1) while the variables turnover and percentage of female workers are statistically significant at α =0.05 and α=0.1 while the constant term is not statistical significant. The model certainly fits the data well and provides evidence that the economical interpretation of the logit model, as the one we tried in this paper, is one of the most useful methods, when a qualitative approach is need to interpretate economical data sets, involving proportions. Similar are the comments in the case of the process

innovations where however the variable Exports seems to have a significant influence both in magnitude as well as in statistical terms.

To assess the model fit we compare the log likelihood statistic (-2 log $\hat{L}$) for the fitted model with the explanatory variables with this value that corresponds to the reduced model (the one only with intercept). The likelihood ratio statistic for comparing the two models is given by the difference

$$LR = -2 \log \hat{L}_R - (-2 \log \hat{L}_F) = 18.121$$

where the subscripts R and F correspond to the Reduced and Full model respectively. The corresponding value of the test in the case of the process innovations is 28.016. That is, in our case the overall significance of the model is X^2=18,121 and 28.016 with a significance level of P=0.000 and 3 and 4 degrees of freedom for the cases of product and process innovations respectively. Based on this value we can reject H_0 (where H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$) and conclude that at least one of the β coefficients is different from zero. These values must be compared with $X^2_{0.05,3}$=7.815 and $X^2_{0.05,4}$=9.488, which implies again a rejection of H_0.

Finally, the Hosmer and Lemeshow values equal to 8.145 and 11.135 (with significance equal to 0.37 and 0.45) for the cases of product and process innovations respectively. The non-significant X^2 value indicates a good model fit in the correspondence of the actual and predicted values of the dependent variable.

Tables 4a and 4b present the estimated odds ratios in the addition of extra variables in the logit formulation and for the cases of product and process innovations respectively. Specifically, the first column presents the variables and the second the odds ratios when running a logit formulation for each variable. That is if we run a logit for the case of turnover the odds ratio equals to 3.833 while if we run a logit only for exports then the odds ratio is 4.778. The other columns refer to the case of running logit models with more than one explanatory variable.

Table 4a: Odds ratio estimated in the case of product innovations

Q2	3.833	2.216	3.219	2.62	3.460	2.963		
Q3	1.434	1.346	1.511				1.176	1.349
Q4	1.148						1.282	
Q6	4.778			3.292				2.201
PFW	1.018		0.973		1.003			
PFU	1.015					1.004		

Table 4b: Odds ratio estimated in the case of process innovations

Q2	2.625	1.932	1.897	2.62	1.936		1.738	
Q3	1.290	1.234	1.271		1.162	1.186		2.696
Q4	1.350					1.331		
Q6	4.444	0.979	3.123	3.292			3.400	1.188
PFW	1.012		0.986	2.62		0.988	1.738	
PFU	1.012							2.696

3.

4. Conclusions and policy implications

In this paper we discuss the Greek case of female participation in innovation activities of firms using the first results of a research project on women in innovation, technology and science. Based on the fitted logistic model the estimated odds ratio equals to

- 3.219 and 1.897 for implementing product and process innovations respectively for firms, which have a higher than 10% increase in turnover. This means that the odds of implementing product and process innovations are about 3.219 and 1.897 times higher respectively for a firm which has an increase of more than 10% of its turnover than for a firm which has not.
- 1.511 and 1.271 for implementing product and process innovations for firms having a higher average life of product. This means that the odds of implementing innovation is about 1.511 and 1.271 times higher for a firm which higher average product life compared with those with low and in the cases of product and process innovations respectively.
- 0.973 and 0.986 for implementing product and process innovations for firms with a high percentage of female workers.
- 3.123 for implementing process innovation for firms, which have exportations.

Similarly, in relation to turnover the odds of implanting innovations increase by 221.9% and 89.7% for product and process innovations respectively ceteris paribus. We can say that in relation to the average life of product the odds of implementing innovations increase by 51.1% and 27.1% ceteris paribus. In relation to percentage of female workers the odds of implementing product and process innovations decreases by 0.03% and 0.015% respectively, all the other remaining fixed in each case. This certainly has as a result the fact that the percentage of the participating acting women does not influence the implementation of the innovation, in the Greek Entrepreneurship.

Finally, the relationships were positive except in the case of the percentage of female workers where we have a negative relationship between implementation of innovations and female workers. This could be explained as a higher percentage of female workers corresponding to a lower implementation of innovations. This can be considered in relation to other Departments in the firm like R&D as well as to the markets that the firm operates.

Greece, in order to develop future capabilities and make the necessary choice for technological priorities, needs a more comprehensive cooperative innovative effort. The most important factors influencing the incidence of innovation and the speed of its diffusion are:

(a) Technical applicability;
(b) Profitability;
(c) Finance,
(d) Size, structure and organisation and
(e) Management attitudes.

Additionally, some other important factors may be needed like R&D, easy access to available information and the labour market availability of certain skills.

ACKNOWLEDGMENTS

We have both participated in the research project entitled «Women and Innovation: Determinant factors and obstacles of innovative activity of Greek firms: 2000-2003» co-funded by 25% from the Greek Government and 75% from the European Union under the framework of the «Education and Initial Vocational Training Program-Archimedes». We would like to thank the Ministry of Education for their economical support.

References

Cox, D.R. and Hinkley, D.V. (1974). Theoretical statistics. Chapman and Hall, London

Foray, D. (2000). "Characterising the Knowledge Base: Available and Missing Indicators", pp. 239-255. In OECD. ed. *Knowledge Management in the Learning Society.* Paris: OECD, 2000.

Hair, J.F, Anderson R.E, Tatham, R.L and Black, W.C. (1998). Multivariate data analysis, Prentice Hall, Fifth Edition

Halkos G. E. (2006). *Econometrics: Theory and Prac*tice, Giourdas Publications

Hosmer D., W and Lemeshow S. (1989). *Applied logistic regression,* John Wiley and Sons, New York

Kitsos C. P. *et.al*. (2006). *Women and Innovation: Determinant factors and obstacles of innovative activity of Greek firms: 2000-2003: ARCHIMEDES research project*, Athens.

Kitsos C. P. and Edler L. (Eds) (1997). *Industrial Statistics*. Physica-Verlag.

Kitsos, C. P. (2006). On the Logit Methods for Ca Problems. In Statistical Methods for Biomedical and Technical Systems, by F. Vonta (Ed), Limassol, Cyprus, pg 335-340.

Kleinbaum D.G. (1994). *Logistic regression: A self learning text*, Springer-Verlag, New York.

Kleinbaum D., G. and Kupper L.,L., Muller KE. and Nizam A. (1999). *Applied regression analysis and other multivariate techniques*, Duxbury, Third Edition.

Myers, S. and Marquis, D.G. (1969) *Successful Industrial Innovation.* Washington D.C.: National Science Foundation.

Nedler, J.A. and Weddenburn, R.W. M. (1972). Generalized Linear Models, *Journal of Royal Statistical Society* A, 135, 370-384.

Nelson, R.R. (1993). *National Innovation Systems: a Comparative Analysis.* Oxford Univ Press. Oxford

OECD (1996) . *The Knowledge-Based Economy.* Paris : OECD, 1996.

OECD (1999). *Managing National Innovation Systems.* Paris : OECD.

OECD (2001). *Innovative Clusters: Drivers of National Innovation Systems.* Paris: OECD.

Scherrish, M.J. (1995). *Theory of Statistics,* Springer Series in Statistics

Sharma S. (1996). *Applied multivariate techniques,* John Wiley and Sons, New York

Schmookler, J. (1996). *Invention and Economic Growth.* Cambridge: Harvard Univ. Press

Von Hippel, E. (1994). *The Sources of Innovation.* Oxford : Oxford University Press.