



Journal of Management Sciences and  
Regional Development  
Issue 5, July 2005  
Correspondence: [ikarkazis@aegean.gr](mailto:ikarkazis@aegean.gr)

<http://www.stl.aegean.gr/geopolab/GEOPOL%20PROFILE.htm>  
ISSN 1107-9819  
Editor-in-Chief: John Karkazis

## THE USE OF SURVIVAL ANALYSIS TECHNIQUES IN STUDYING REGIONAL DEVELOPMENT ISSUES.

**V. Angelis**

*Department of Business Administration  
University of the Aegean  
Chios, Greece*

**C. Dimaki and I. Virras**

*Department of Statistics  
Athens University of Economics and Business  
Athens, Greece*

**Abstract.** The state of development of a given region at a point in time may be expressed through its "Image", an index describing the region's economic and social conditions. Regions grow and decline. In this sense regions may be considered as «patients» whose «health» status is given by the values of their Image. Treatment is defined as the set of actions taken, by the central or the local authorities, to improve a region's Image. Those actions include a number of financial incentives such as grants, tax relieves, subsidised loans etc. A region is considered as a survivor as long as its Image follows an increasing trend or at least remains constant. The moment of change in trend direction from increasing to decreasing indicates the failure (death) of this region.

The situation described above indicates that Survival Analysis, a collection of statistical procedures of data analysis for which the outcome variable of interest is time until an event occurs, may be an appropriate tool to use in studying regional development issues. This paper consists of two parts. The first part covers the theoretical aspect of the subject by introducing the concept of an area's Image as well as a number of Survival Analysis procedures. The second part applies those procedures on data drawn from the 51 counties of Greece in order to test their effectiveness in measuring the effect of a Regional Development Act on their development.

**Keywords:** Economic Development, Image of a Region, Regional Development, Social Development, Survival Analysis

## **1. INTRODUCTION**

The growth or decline of a region depends on its power to attract both industries and the right blend of people to run them. This attractiveness depends on what we may call the Image of the area (Angelis, 1981). At each point in time the region «sends out» its Image and depending on its impact on the people (both employers and employees) the area may be considered as Attractive or Repulsive. The Image of an area, however, is not fixed but changes over time. Its variations depend both on the area's internal momentum and external interventions. Furthermore those variations may be of two kinds: variations where an area retains its present status (attractive or repulsive) and variations where the area's status changes.

An external intervention of particular importance is state intervention. This may take different forms (infrastructure improvement, provision of financial incentives to industries such as tax-reduction, low-interest loans, grants, subsidies etc.), may have varying levels of intensity and may be applied to various regions so as to improve their Image values and hence their development prospects.

The objective of this paper is to use the Survival Analysis techniques in order to test the effect of financial incentives on region's development. Regions are considered as "patients" whose «health» status is given by the values of their Image. Treatment is defined as the set of actions taken, by the central or the local authorities, to improve a region's Image. Those actions include a number of financial incentives such as grants, tax relieves, subsidised loans etc. A region is considered as a survivor as long as its Image follows an increasing trend or at least

remains constant. The moment of change in trend direction from increasing to decreasing indicates the failure (death) of this region.

After this brief introduction, Section 2 introduces the concept of an area's Image, Section 3 presents the basic concepts of Survival Analysis and its main techniques (Crowder et al (1995), Lee (1992)), Section 4 applies those techniques on data drawn from the 51 counties of Greece in order to test the effect of a Regional Development Act on their development (Angelis & Dimaki (1998), Virras (1999)) and finally Section 5 summarizes the conclusions and suggest areas for further research.

## **2. THE CONCEPT OF AN AREA'S IMAGE**

As it has been mentioned already the attraction power of an area is expressed by what we may call the Image of this area. However, one may argue that since people "receiving" the Image of an area belong to various distinct groups (i.e. employers, professionals, unskilled workers, skilled workers, etc) and are sensitive to different factors, the impact of the area's Image on the members of each particular group will be different. Whilst this is plausible, empirical evidence suggests that all groups of potential movers react similarly to a basic set of factors; more precisely, a set of minimum standards, largely common to all groups, must be satisfied if the area is to be considered as a potential choice by any of them. To reconcile these two views we refine the concept of an area's Image by introducing the following two concepts:

Basic Image and Specific Image

The concepts of Basic and Specific Images have been discussed in full detail in some earlier papers and the most important points are summarized below (Angelis (1981), (1990), Angelis & Dimopoulou (1991)).

The Basic Image of a given area measures the degree to which this area satisfies a set of basic criteria common for all movers. An area satisfying those criteria is considered, by all potential movers, as worth a closer examination and as a potential final choice. On the other hand, the Specific Image of a given area as perceived by a particular group of potential movers measures the degree to which movers belonging to that particular group consider this area as their final choice.

The Basic Image of an area may be expressed as a multitude of factors (Cullingworth (1969), Hunter & Reid (1968), Rhodes & Khan (1971), Rostow (1960), Townroe (1971)). Those factors may be divided into two groups, depending on whether they refer to its economic or its social function. The factors of the first group (Accessibility to Materials and Markets, Land Availability, Financial Conditions) give a measure of the actual economic and industrial potential of the area, called Economic Indicator. Similarly the factors of the second group (Housing Conditions, Environmental Conditions, Social Conditions) give an estimate of the actual social conditions in the area, called the Social Indicator. Therefore, we could say that the Basic Image (*BI*) of an area may be expressed as a function of two conflicting indicators, Economic (*EI*) and Social (*SI*). Hence,

$$BI = f(EI, SI)$$

Furthermore, there are indications that this function is non linear and its graph is discontinuous. In order to study this function a mathematical method of looking at discontinuous phenomena,

developed by Rene' Thom, was used (Isnard & Zeeman (1976), Thom (1975), Zeeman (1973)). This method is called Catastrophe Theory and is particularly applicable in cases where continuous causes have discontinuous effects. According to this theory, the value  $x$  of an area's Basic Image at every period of time is given by the equation:

$$i^3 - Bi - A = 0 \quad (2.1)$$

where  $A, B$  are functions of the area's Economic and Social Indicator.

Concluding we note that the Basic Image value of any given area lies in the interval  $[-1, 1]$ . Positive Basic Image indicates an area that may be considered as a potential final choice by the various groups of prospective movers.

The Specific Images of an area as perceived by the various groups of movers are primarily influenced by the area's Basic Image. Additionally however each group of movers is also influenced by several other factors specific to this particular group. In the case of investors the most important of those factors is the provision of financial incentives. Hence,

$$SPII = g(BI, FINI)$$

where  $SPII$  is the area's Specific Image as perceived by potential investors and  $FINI$  is a function of the Financial Incentives provided to the investors when moving into this area. For the purpose of this work the Specific Image values of any given area lie in the interval  $[0, 2]$ .

Obviously, Financial Incentives are used as a means of "improving" an area's Specific Image and "pushing" inventors into it, hoping that this growth generated will eventually lead to self-sustained development. However, experience has shown that the effects of the financial incentives on the development of an area are weak and

temporary, unless they are accompanied by measures aiming at improving the area's Basic Image.

By keeping the Basic Image of a city attractive, we make sure that, in spite of any possible fluctuations in the effectiveness of various specific factors and unexpected external adversities, the area may retain its overall pulling power, renew its ageing industries, maintain the right blend of workforce and finally overcome any difficulties. As soon as the Basic Image becomes repulsive, however, the situation changes completely; the city enters a vicious circle of deprivation and decline the breaking of which is extremely difficult. Piecemeal approaches, aiming at the breaking of this vicious circle, through the improvement of certain specific factors, may help temporarily but the only lasting solution to this problem is the restoration of the Basic Image.

### **3. SURVIVAL ANALYSIS TECHNIQUES**

#### **3.1. Basic Concepts**

Survival Analysis is a collection of statistical procedures of data analysis for which the outcome variable of interest is time until an event occurs (Cox & Oakes (1992), Crowder et al (1995), Kalbfleisch & Prentice (1980), Lee (1992), Parmar & Machin (1995)). By time, we mean the period from the beginning of the follow-up of an individual until an event occurs; alternatively, time may refer to the age of an individual when an event occurs. We usually refer to the time variable as survival time, because it gives the time that an individual has "survived" over some follow-up period.

We denote by  $T$  ( $T \geq 0$ ) the random variable for a person's survival time and by  $t$  any specific value of interest for the random variable  $T$ . By event, we mean any designated experience of interest that may happen to an individual. We usually refer to the event as a failure, because in most of the cases it is a negative experience. However an event may also be a positive experience in which case survival time will be the period up to that event. Most survival analysis cases face a key analytical problem called censoring. Censoring occurs when we have some information about an individual's survival time, but we don't know the survival time exactly. Censoring may in general occur in the following three cases:

- A person does not experience the event before the study ends;
- A person is lost to follow-up during the study period;
- A person withdraws from the study.

### **3.2 Survival and Hazard Functions**

Having presented the key concepts, we are now ready to introduce and describe two quantitative terms considered in any survival analysis situation. These are the survivor function, denoted by  $S(t)$ , and the hazard function, denoted by  $h(t)$ .

The survivor function  $S(t)$  is defined for both discrete and continuous random variables as the probability that a person survives at least as long as some specified time  $t$ , i.e.  $S(t) = P(T \geq t)$ ,  $0 < t < \infty$ . Theoretically, as  $t$  ranges from 0 up to infinity, the survivor function can be graphed as a smooth curve. All survivor functions have the following characteristics:

- they are monotone nonincreasing
- at  $t = 0$ ,  $S(t) = S(0) = 1$
- as  $t \rightarrow \infty$ ,  $S(t) \rightarrow 0$

Note that these are theoretical properties of a survivor curve. In practice we usually obtain graphs that are step functions, rather than smooth curves.

The hazard function  $h(t)$  gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time  $t$ . Note that, in contrast to the survivor function, which focuses on not failing, the hazard function focuses on failing, that is, on the event occurring. The hazard function  $h(t)$  is most easily specified separately for discrete and continuous  $T$ .

In the case where  $T$  is continuous

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$= - \left[ \frac{dS(t)/dt}{S(t)} \right], t \in [0, \infty]$$

Similarly, in the case where  $T$  is discrete

$$h(t) = P(T = t | T \geq t)$$

$$= \frac{P(T = t)}{S(t)}, t = 0, 1, \dots$$

Regardless of which function  $S(t)$  or  $h(t)$  one refers to there is a clearly defined relationship between the two. In fact, if one knows the

form of  $S(t)$ , he can derive the corresponding  $h(t)$ . It is well known that  $S(t)$  is a common representation of the distribution of a random variable. The hazard function is a more specialized characterization, particularly useful in modeling survival time data, since in many cases information is available as to how the failure rate will change with the length of time on test. The information can be used to model  $h(t)$  and it is easily translated into information for  $S(t)$ , since it can be proved that  $S(t)$  is uniquely determined by the form of  $h(t)$ . Indeed in the case where  $T$  is continuous

$$S(t) = \exp \left[ - \int_0^t h(u) du \right]$$

and in the case where  $T$  is discrete

$$S(t) = \prod_{i=1}^{t-1} [1 - h(i)].$$

### 3.3 Nonparametric Methods for Estimating and Comparing Survival Functions

After this short introduction into the basic concepts of Survival Analysis we will now go to outline the major methods of estimating and comparing survival functions.

#### 3.3.1 *The Product-Limit Method of estimating Survival Curves of individuals or items that belong to different groups.*

Suppose that there are  $k$  distinct times  $t_{(1)} < t_{(2)} < t_{(3)} < \dots < t_{(k)}$  at The *Product-Limit* or *Kaplan-Meier* method of estimating Survival Curves is a simple and easily applicable non-parametric technique, which can provide us with useful information regarding the probability

of an item or of an individual to survive longer than time  $t$ . It is very useful when we are interested to study the effect of a certain factor upon the survival probabilities which failures occur. Let  $d_j$  be the number of failures at time  $t_j$ . Let  $r_j$  be the number of units at risk at time  $t_j$ , which means that  $r_j$  is the number of units that haven't failed or are not censored prior to  $t_j$ . Then, the product-limit estimate of  $S(t)$  is given by,

$$\hat{S}(t) = \prod^{(t)} \left( 1 - \frac{d_j}{r_j} \right)$$

where by  $\prod^{(t)}$  we denote the product over all  $j$ 's is that satisfy the condition  $t_{(j)} < t$ . The plot of  $\hat{S}(t)$  versus  $t$  is a step function, known as **Kaplan – Meier curve**, which is very informative, particularly when we compare the lifetime of two or more groups of units, which differ with respect to an attribute.

### 3.3.2 Non-parametric Methods for comparing Survival Functions

In the statistical literature one can find a large number of non-parametric tests for the comparison of two or more survival curves. For each test we suppose that we have two groups of units differing with respect to one factor, whose effect on the survival probability we want to study. This effect is studied through the comparison of the respective survival curves, by appropriate statistical tests. The hypotheses we are interested to test, might be that the two survival curves simply differ (two-sided test), or that one survival curve is superior to the other, which would mean that the units of the respective group have greater probability to survive than the units of the other group (one-sided test). The most important of those non-parametric tests are presented below.

### 3.3.2.1 The Cox – Mantel Test

Suppose that we have two groups of units with Survival Curves  $S_1$  and  $S_2$  respectively. The units of the first group have some typical characteristics representing the normal conditions while the units of the second group differ as compared to those of the first with respect to a certain feature the effect of which on the survival of the units we want to study. We combine the failure and censored times of the units of both groups into a new one and we rank them in ascending order. We denote by  $t_{(1)} < t_{(2)} < \dots < t_{(n)}$  the distinct failure or censored times of the new group and by  $d_j$  the number of failures that took place till the time  $t_j$ . The set  $R(t_j)$  represents the units that are exposed to risk failure prior to time  $t_j$ , i.e. the number of the units that are not censored or they haven't failed prior to  $t_j$ . Let  $r_{1j}$  and  $r_{2j}$  be the number of units that are exposed to risk failure  $R(t_j)$  and belong to the first and to the second group respectively. The total number of units in the risk set at each time  $t_j$  is given by the sum  $r_{(j)} = r_{1j} + r_{2j}$ . The test statistic is given by the ratio  $C$

$$= \frac{U}{\sqrt{I}} \text{ where the quantities } U \text{ and } I \text{ are defined as } U = r_2 - \sum_{j=1}^k d_j A_j$$

$$\text{and } I = \sum_{j=1}^k \frac{d_j (r_{(j)} - d_j)}{r_{(j)} - 1} A_j (1 - A_j) \text{ respectively and } A_j \text{ is the proportion}$$

of  $r_{2j}$  into the total  $r_j$ . It can be proved that the test statistic  $C = \frac{U}{\sqrt{I}}$  is distributed asymptotically as a standard normal variable under the null hypothesis  $H_0 : S_1 = S_2$ .

### ***3.3.2.2 The Logrank Test***

This test uses mainly the observations of one of the two groups in order to draw conclusions about the differences in the survival probability between the two groups. The main procedure of the test is to assign to each observation of both groups a score which is a function of the logarithm of the of the respective group's Survival function and differs in the general form according to whether the observation is censored or not. The test statistic for the Logrank test is based on the sum of the scores of one of the two groups and the respective variance of the sum of the scores and has an asymptotically standard Normal distribution under the null hypothesis.

### ***3.3.2.3 Peto and Peto's Generalised Wilcoxon Test***

The Peto and Peto's Generalised Wilcoxon Test is similar to the Logrank test. As in the case of the Logrank, this test is based on assigning a score to each observation in both groups, taking into account whether this observation is censored or not. This specific test uses the sum of the scores of one of the two groups and its respective variance, in order to compare the survival curves of the two groups and to draw conclusions about the difference in the survival probability between the groups, which are due to the effect of the certain factor differing between the two groups. The whole procedure of the test is the same as in the case of the Logrank test. First we calculate the scores for each observation for both groups; the censored observations have negative scores while the uncensored observations have positive scores. The total sum of the scores is equal to zero as in the case of the Logrank test. Then we calculate the sum of the scores for one of the two groups and its

respective variance, which is the same as in the case of the Logrank test. The test statistic used has an asymptotically standard Normal distribution under the null hypothesis.

### 3.3.2.4 The Gehan's Generalised Wilcoxon Test

The realization of the specific test is based on comparisons between each observation of the first group against each observation of the second group. At each comparison we assign an appropriate score, which is related to result of the comparison. If we have two groups with  $n_1$  and  $n_2$  observations respectively, then in order to carry out the specific test we must realize  $n_1 n_2$  comparisons in total and assign the respective scores to each one. The scores used in this specific test depend on the kind of the hypothesis. The procedure is easy to be carried out when the number of the observations is small enough; otherwise it becomes laborious.

Mantel proposed an alternative method for the realization of this test and has also shown that this method is equivalent to the previous one. The general idea of this method is that instead of comparing the observations of the two groups, one can construct a combined sample consisting of the observations of both samples and assign to each observation, scores related to the relative ranking of each observation in the combined sample. In this way we compare each observation  $i$  with the rest  $n_1 + n_2 - 1$  observations. The test statistic is  $W = \sum_{i=1}^{n_1} U_i$ . Under the null hypothesis the sum of the scores  $W$  has an asymptotically Normal

$$\text{distribution with mean zero and variance } Var(W) = \frac{n_1 n_2 \sum_{i=1}^{n_1+n_2} U_i^2}{(n_1 + n_2)(n_1 + n_2 - 1)}$$

### 3.4 Regression Models for Survival Data

In this Section we introduce two methods for identifying risk factors, i.e. factors related to survival time. Those risk factors may be seen as independent variables, quantitative or qualitative, affecting the dependent variable, survival time or probability. The first method, the Proportional Hazard, is applicable when the dependent variable is continuous while the second method, the Logistic Regression, when the dependent variable is binary (dichotomous).

#### 3.4.1 *The Proportional Hazard Model*

When the relationship between the survival time and  $k$  prognostic variables is under investigation, multiple techniques are more efficient than univariate ones, since they take into account simultaneous effects of the variables. The Proportional Hazard model uses the hazard rate function as the dependent variable in order to describe how the various factors affect the survival probability of the unit. In other words through such a model we can investigate whether some factors accelerate or not the failure time of the unit. On the other hand the positive influence of the factor upon the lifetime of the units, is interpreted as longer lifetime.

The general form of a *Proportional Hazard model* is  $h(t; z) = h_0(t)\psi(z)$ . Obviously the hazard depends on both time and covariates, but through two separate factors which may involve unknown parameters. The first,  $h_0(t)$ , can be thought as a baseline hazard function, it is a function of time only, it is usually unknown and corresponds somehow to the conditions for the covariates that are

considered as normal or typical. The second,  $\psi(z)$ , is some positive function of the covariates, used to express the influence of the factors upon the hazard rate.

A special form of the proportional hazards model, when survival times are continuously distributed and the possibility of ties can be ignored, is

$$h(t; z) = h_0(t) \exp(\beta z)$$

$$= h_0(t) \exp\left(\sum_{i=1}^k \beta_i z_i\right)$$

where  $h_0(t)$  is the hazard function of the underlying survival distribution when all the  $z$  variables are ignored and the  $\beta$ 's are regression coefficients. It can be shown that the above mentioned model is equivalent to

$$S(t) = \left(S_0(t)\right)^{\exp\left(\sum_{i=1}^k \beta_i z_i\right)}$$

where  $S_0(t)$ , can be thought as the Survival baseline function.

The parameters  $\beta$  of the function  $\psi(z)$  can be estimated with the use of the Marginal Likelihood as illustrated by Kalbfleisch and Prentice (1973) and the Partial Likelihood as illustrated by Cox (1972), giving alternative solutions, when the data include tied or censored observations.

### 3.4.2 Logistic Regression

In Statistics the term binary data refers to data that come from a random experiment with binary response. We usually denote this response with the set  $\{0,1\}$ , where [0] stands for the expression of the «failure» in the random experiment, and [1] for the «success» in the same random experiment. Let next consider the random variable  $Y = \sum_{i=1}^n Z_i$ , where  $Z_i=1$  if the outcome is a success and  $Z_i = 0$  if the outcome is a failure.

The statistical model that is used to describe such experiments is the binomial distribution, with probability function  $Pr(Y=y) = \binom{n}{y} p^y (1-p)^{n-y}$ ,  $y = 0,1,2,\dots,n$ , where  $p$  is the probability of a success,  $n$  is the total number of trials and  $y$  the desired number of successes. The mean and the variance of the Binomial distribution are given by  $E(X) = np$   $Var(X) = np(1-p)$  respectively. The binomial distribution belongs to the exponential family of distributions since it can be written in the form  $\exp\left\{y \log\left(\frac{p}{1-p}\right) + n \log(1-p) + \log\binom{n}{y}\right\}$ .

The log-likelihood function of  $N$  independent identically distributed Binomial random variables is given by the formula

$$l(p_1, p_2, \dots, p_N; y_1, y_2, \dots, y_N) = \sum_{i=1}^N \left[ y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i \log(1-p_i) + \log\binom{n_i}{y_i} \right].$$

It is natural to assume that to each realisation of a Binomial experiment  $Y_i$ ,  $i=1,2,\dots,N$ , there are assigned covariates-factors that affect the probability of success. These covariates can be either qualitative, such as dummy variables, or quantitative, such as the log-

dose of a substance. As previously these covariates stand for the systematic component of the model, which is of the form  $n_i = \sum_{j=1}^p x_{ij} \beta_j$ ,  $i=1,2,\dots,n$ , where  $\underline{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  is a vector of unknown coefficients.

In order to estimate the vector of the unknown coefficients a Newton-Raphson algorithm is used. The log-likelihood ratio test is used to assess the goodness of fit of the model.

#### **4. AN APPLICATION USING SURVIVAL ANALYSIS TECHNIQUES**

##### **4.1 Introduction**

In this section we will use the Survival Analysis Techniques presented so far to test their effectiveness in measuring the effect of Development Act 1262/82 on Regional Development (Angelis & Dimaki (1998), Virras (1999)). For this purpose we consider:

- As patients the 51 counties of Greece.
  - As follow-up period of their health status the period 1971-1993.
- The choice of this period has been dictated by the fact that during this time the most important Regional Development Act (Act 1262/1982) has been introduced by the Greek authorities in order to assist selected underdeveloped regions. The data resulting from this Act's implementation provide a suitable basis for our analysis. Although

extending the series and rerunning the model is within our plans the results obtained with the present series are considered quite reliable.

- As treatment given to improve their health status the financial incentives provided by the Act 1262/1982. This Act divides the counties into 4 groups (Table 1) and the strength of the incentives provided to them increases as we move from Group 1 to Group 4. The aim of these incentives is to rapidly and temporarily improve the area's Specific Image as perceived by industries thus "pulling" inventors into it and hope that the growth generated in that way will eventually lead to a self-sustained growth and a permanent improvement of the area's both Basic and Specific Images.

Primary data have been collected for those counties for the period 1971-1993 and their respective Basic Images have been calculated for each year of this period using the method presented in the first part of this work. Furthermore the Basic Image values of every county for the period 1971-1993 have been considered as time-series and the Hodrick-Prescott Filter (Hodrick and Prescott, 1997, Pedersen, 1999) has been used to estimate and plot their trend.

It is reminded that within the framework of this work a region is considered as a «**survivor**» as long as its basic Image follows an increasing trend or at least remains constant. A change in trend direction from increasing to decreasing indicates the **death** of this region. Furthermore as «**success**» we define the event that a county has a positive Basic Image while as «**failure**» the event that a county has a negative Basic Image.

GROUP 1	GROUP 2	GROUP 3	GROUP 4
ATTICA	ACHAIA	AITOLOAKARNANIA	DODEKANISA
THESSALONIKI	VOIOTIA	ARGOLIDA	DRAMA
	HERAKLEIO	ARKADIA	EVROS
	KORINTHIA	ARTA	FLORINA
	LARISA	EYVOIA	IOANNINA
	MAGNISIA	EYRITANIA	KASTORIA
		FOKIDA	KERKYRA
		FTHIOTIDA	KILKIS
		GREVENA	LESVOS
		ILEIA	MESSINIA
		IMATHIA	PELLA
		KAVALA	RODOPI
		KARDITSA	SAMOS
		KEFALLONIA	SERRES
		KOZANI	THESPROTIA
		KYKLADES	XANTHI
		LAKONIA	CHIOS
		LASITHIOU	
		LEYKADA	
		PIERIA	
		PREVEZA	
		RETHYMNO	
		TRIKALA	
		CHALKIDIKI	
		CHANIA	
		ZAKYNTHOS	

**Table 1: Classification of Counties according to Act 1262/82**

Looking back at the actual development process of the Greek counties during the period under study and especially after the Act's implementation we may note the following points:

- A steady state development for the developed counties (Group 1), which were given practically no incentives.
- A slight improvement for the mildly undeveloped counties (Group 2 and 3), which received considerable assistance.
- A stagnation / slight improvement for the less developed counties (Group 4). Furthermore within that group the isolated counties exhibit the worse performance and among them the islands seem to have the bigger problem.

The remaining part of this section examines the extent to which the various Survival Analysis techniques presented so far may trace the difference in the trend of Basic Image values for counties belonging to various groups or possessing certain characteristics.

## **4.2 The Product-Limit Method**

Our objective in this section is to test the effect of a county's spatial continuity on its development. Hence, the 51 counties of Greece are divided, on the basis of their spatial continuity, into two groups, island and non-island ones (Table 2), and the Product-Limit Method is used to examine if the location of a county affects its survival time i.e. the time that its Basic Image trend remains increasing or at least constant.

ISLAND COUNTIES	NON - ISLAND COUNTIES		
CHANIA	ACHAIA	FTHIOTIDA	MAGNISIA
CHIOS	AITOLOAKARNANIA	GREVENA	MESSINIA
DODEKANISA	ARGOLIDA	ILEIA	PELLA
HERAKLEIO	ARKADIA	IMATHIA	PIERIA
KEFALLONIA	ARTA	IOANNINA	PREVEZA
KERKYRA	ATTICA	KARDITSA	RODOPI
KYKLADES	CHALKIDIKI	KASTORIA	SERRES
LASITHIOU	DRAMA	KAVALA	THESSPROTIA
LESVOS	EVROS	KILKIS	THESSALONIKI
LEYKADA	EYRITANIA	KORINTHIA	TRIKALA
RETHYMNO	EYVOIA	KOZANI	VOIOTIA
SAMOS	FLORINA	LAKONIA	XANTHI
ZAKYNTHOS	FOKIDA	LARISA	

**Table 2: Classification of the Greek Counties according to their spatial continuity**

Figure 1 presents the Survival Curves as well as the Cumulative Survival Proportion for each group of counties.

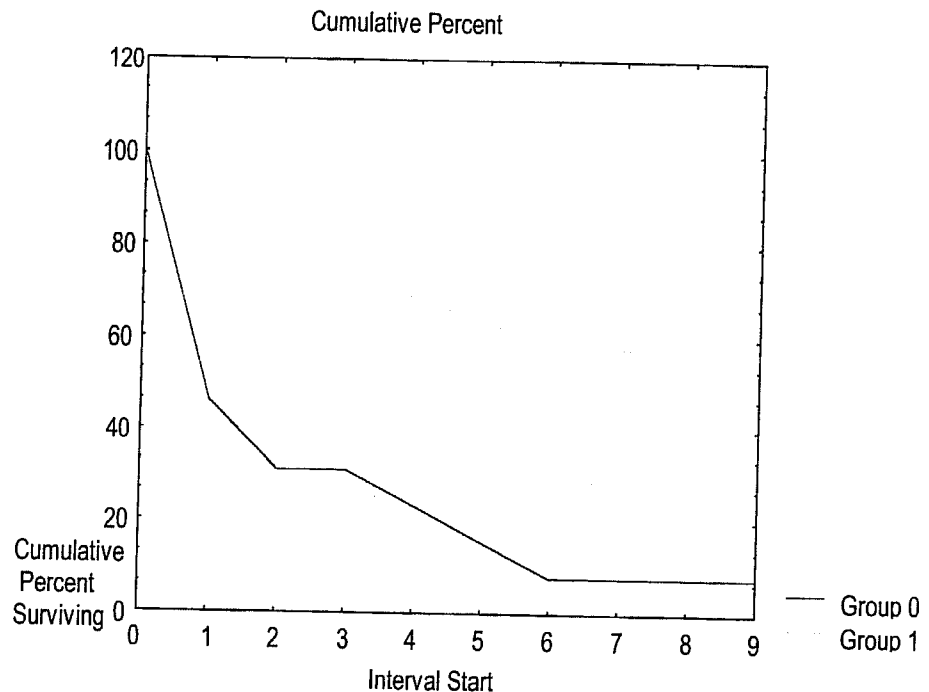
*Cumulative Survival Proportion for Both Groups***Figure 1 Comparison of the Survival Curves**

Table 3 summarizes the results obtained by using the appropriate non-parametric tests i.e. Gehan's Generalised Wilcoxon Test, Cox-Mantel Test, Log-Rank Test and Peto's and Peto's Generalised Wilcoxon Test for comparing the two curves.

Looking at both the curves of Figure 1 and the tests' results of Table 3 we conclude that there is strong evidence to suggest that, at significance level  $\alpha=0.05$ , the survival curve of the second group (non-island regions) is higher than the respective curve of the first group (island regions). This effectively means that the Basic Image trend of the regions belonging to the second group remains increasing or at least

constant, during the period under study, longer than the respective trend of the counties belonging to the first group.

This is an expected conclusion. Island regions have a comparative disadvantage over the rest due to high transportation costs but mainly due to accessibility difficulties and the subsequent feeling of isolation. As a result their power in attracting new investments is limited as compared to that of most of the non-island regions. All measures taken to improve this situation aimed at « pushing » business units into the island regions through the provision of mainly short-term financial incentives. Obviously, the effect of such measures, if they are not taken as part of well designed long term regional policy, is very limited. This is exactly expressed by our conclusions. Sporadic and temporary improvements don't change the overall negative picture in island regions.

1. Gehan's Generalised Wilcoxon Test			
W	Var(W)	Z	p-value
301	7685.5	3.43345	0.0003

2. The Cox - Mantel Test			
U	I	C	p-value
7.24602	3.284848	3.99799	0.00003

3. The Log-Rank Test			
S	Var(S)	L	p-value
-7.246	4.7038	-3.340975	0.00042

4. The Peto & Peto's Generalised Wilcoxon Test			
S	Var(S)	Z	p-value
-5.9020	2.9548	-3.433453	0.0003

**Table 3: Nonparametric tests for comparing survival distributions**

$H_0: S_0=S_1$  vs  $H_1: S_0<S_1$

### 4.3 The Proportional Hazard Model

This is, in some sense, a more detailed analysis of the one presented in the previous section. What we are looking for now are the factors affecting the survival time of a region. The factors considered are the location of the region and the level of treatment (subsidy) the region receives from the state. On the basis of the first factor the 51 counties of Greece are classified into two groups; island and non-island regions (Table 2).

On the basis of the second factor those counties are classified into four groups (Table 1). The level of treatment for each group is taken to be the mean of the percentage subsidy received by each group of counties during the period under study (Act 1262/1982). Hence the values are 0.175, 0.175, 0.270 and 0.350 respectively for the four groups. However, since the first two of the four subsidy groups contain a small number of observations the analysis will be finally limited to the last two groups).

Model	Log-likelihood	Parameters	-2logR( $\theta$ )	p-value
Saturated	-81.5731	4	-	-
Additive	-81.7469	3	0.3476	0.555
Treatment	-86.3346	2	9.1754	0.002
Group	-82.1128	2	0.7318	0.392
Constant	-86.4638	1	8.7020	0.003

**Table 4: Proportional Hazard Model, Table of Log-Likelihood**

Parameter	Beta	Stand. Error	t-value	Exponent Beta
Group	-1.28873	.417144	-3.08942	.275620

**Table 4A: Estimated Parameters for Group Model**

Looking at the results obtained from the application of the Proportional Hazard Model (Tables 4, 4A) we may draw the following conclusions:

- ◆ At a significance level of  $\alpha=0.05$  the non - island regions display a better development trend than the island regions
- ◆ At the same significance level the level of subsidy received doesn't seem to have any considerable effect on the development pattern of the regions.

Referring to the first conclusion we may say that it completely verifies the conclusion drawn in the previous section. The effect of the measures taken to increase the attractiveness of island regions is very limited and sporadic and the temporary improvements don't change the overall negative picture in island regions.

Referring to the second conclusion we may say that the differences in the levels of subsidies, provided to industries moving into the two groups of counties under study, didn't have any statistically significant impact in each group. This is probably due to the fact that the levels of subsidies are similar in those two groups.

#### 4.4 Logistic Regression

In this last section we are looking for the factors affecting the sign of a region's Basic Image. The factors considered are the spatial continuity of the region and the centrality of the region's location.

On the basis of the first factor the 51 counties of Greece are classified into two groups: island and non-island regions (Table 2). On the basis of the second factor those counties are also classified into two groups, border and non-border regions (Table 5).

BORDER DEGIIONS		NON BORDER REGIONS	
CHIOS	ACHAIA	KARDITSA	TRIKALA
DODEKANISA	AITOLOAKARNANIA	KAVALA	VOIOTIA
DRAMA	ARGOLIDA	KEFALLONIA	ZAKYNTHOS
EVROS	ARKADIA	KORINTHIA	
FLORINA	ARTA	KOZANI	
IOANNINA	ATTICA	KYKLADES	
KASTORIA	CHALKIDIKI	LAKONIA	
KERKYRA	CHANIA	LARISA	
KILKIS	EYRITANIA	LASITHIOU	
LESVOS	EYVOIA	LEYKADA	
PELLA	FOKIDA	MAGNISIA	
RODOPI	FTHIOTIDA	MESSINIA	
SAMOS	GREVENA	PIERIA	
SERRES	HERAKLEIO	PREVEZA	
THESPROTIA	ILEIA	RETHYMNO	
XANTHI	IMATHIA	THESSALONIKI	

**Table 5: Classification of the Greek Counties according to the centrality of their location**

The Logistic Regression Method was applied to two sets of data, the first referring to the year 1984 (i.e. very shortly after the implementation of the Act under consideration) and the second to the year 1993 (i.e. almost a decade after the Act's implementation).

Looking at the results obtained from the application of the Logistic Regression Method to the two sets of data (Tables 6, 6A, 7, 7A) we may draw the following conclusions.

At a significance level of  $\alpha=0.10$  for the 1984 data set and at a significance level of  $\alpha=0.05$  for the 1993 data set only the spatial discontinuity factor has an impact on the sign of Basic Image. In fact it is effectively impossible for an island region to have positive Basic Image.

Model	-2*Log-likelihood	Parameters	-2logR( $\theta$ )	p-value
Saturated	52.95804	4	-	-
Additive	53.33961	3	0.38157	0.5368
Border	57.06052	2	3.72091	0.0537
Island	55.87500	2	2.53539	0.1113
Constant	59.94468	1	4.06968	0.0437

**Table 6: Logistic Regression Model – 1984. Table of -2\*Log-Likelihood**

Parameter	Beta	Exponent Beta
Constant	-2.48491	0.0833
Island	1.830980	6.2399

**Table 6A: Parameter Estimates for Island Model**

Referring to the above mentioned conclusions we may say that the high transportation cost, but mainly the accessibility difficulties and

the feeling of isolation limit the attractiveness of island regions and keep their Basic Images negative. The same but to a considerable lesser extend happens to the border regions.

Model	-2*Log-likelihood	Parameters	-2logR(θ)	p-value
Saturated	56.01521	4	-	-
Additive	56.54137	3	0.52616	0.4682
Border	61.62230	2	5.08093	0.0242
Island	58.03323	2	1.49186	0.2219
Constant	63.44901	1	5.41578	0.0200

**Table 7: Logistic Regression Model – 1993. Table of -2\*Log-Likelihood**

Parameter	Beta	exponent Beta
Constant	-2.48491	0.0833
Island	2.057463	7.8260

**Table 7A: Parameter Estimates for Island Model**

## 5. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Regions grow and decline over time and in this sense they may be considered as "patients" whose "health" status is given by the values of their Basic Image. Treatment is defined as the set of actions taken, by the central or the local authorities, to improve a region's Basic Image. Those actions include a number of financial incentives such as grants, tax relieves, subsidized loans etc. A region is considered as a "survivor"

as long as its Basic Image follows an increasing trend or at least remains constant. The moment of change in trend direction from increasing to decreasing indicates the "failure" (death) of this region. This approach indicates that Survival Analysis may be an appropriate tool to use in studying regional development issues.

Our objective in this paper was to present certain Survival Analysis techniques and examine their applicability in the field of regional development. Those techniques are then applied on data drawn from the 51 counties of Greece in order to test the effect of Specific Regional Development Laws in their development.

Returning to the Survival Analysis techniques and to the extent to which they can trace the real trend of Basic Image values for the various counties after the implementation of this Act we may note the following:

- The application of Product-Limit method shows that spatial discontinuity is a factor hindering a region's development. Hence, island regions exhibit an overall worse performance than the non-island ones.
- The application of the Proportional Hazard model verifies the above comparison. It shows that the incentives provided to the various counties have a slight positive overall effect on the development of the regions belonging to Groups 3 and 4. However the benefits were much less for the island regions.
- The application of Logistic Regression verifies the previous conclusions. It shows that spatial discontinuity is a key factor affecting the sign of a region's Basic Image. This effectively means that it's practically impossible for an island region to have a positive

Image in spite of the strong financial incentives offered to investors in order to move there.

Summarizing we may say that all three methods presented are successful in tracing the changes in a county's Image as a result of the financial incentives provided by the State for incoming business units.

The analysis presented above may be extended in various directions. Suggested areas for further research include:

- Application of a Bayesian approach of the Cox-Regression on the available data, using the BUGS code and comparison of the results obtained by the two approaches; Classical and Bayesian.
- Study of Regional Development using a Multivariable Time Series Analysis on the Basic Image Index of the various regions and investigation for possible existence of cycles on the series.
- Study of Regional Development using Spatial Analysis.

## REFERENCES

1. Angelis, V (1981). Basic Image. A New Approach in Modeling Growth and Decline in Industrial Cities. *Proceedings Volume, Conference on Dynamic Non-Linear Theory and General Urban/Regional Systems*, Organized by the U.S. National Science Foundation and the U.S. Department of Transportation, Washington D.C. USA, pp. 91-131. (Invited Paper).
2. Angelis, V. (1990). Factors affecting the movement and location of both industrial units and their employees, *Proceedings Volume, J. Karkazis and J. B. Boffey eds., Meeting IV of the Euro Working Group on Location Analysis*, Chios, pp. 219-236.
3. Angelis, V. and Dimaki, C. (1998). Changes in the Attractiveness of a Geographical Region: A Survival Analysis Approach. *Studies in Regional and Urban Planning*. Issue 6, 19-34.
4. Angelis, V. and Dimopoulou, M. (1991). A Decision Support System for the Location of Industrial Units. Theoretical Framework and Applications. *Studies in Locational Analysis*, 3, pp. 83-99
5. Cox, D.R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society*, 34, 187-220.
6. Cox, D.R. and Oakes, D. (1992). *Analysis of Survival Data*. Chapman & Hall, London.
7. Crowder, M.J. et al (1995). *Statistical Analysis of Reliability Data*. Chapman & Hall, London.
8. Cullingworth, J.B. (1969). *Housing and Labour Mobility*, OECD, Paris.
9. Hodrick, R. & Prescott, E. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money Credit & Banking*, 1-16.

10. Hunter, L. and Reid, G. (1968). *Urban Worker Mobility*, OECD, Paris.
11. Isnard, C.A. and Zeeman, E. C. (1976). Some Models from Catastrophe Theory in the Social Sciences in Collins, L. (ed.) *Use of Models in the Social Sciences*. Tavistock Press, London.
12. Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*, Wiley, New York.
13. Lee, E.T. (1992). *Statistical Methods for Survival Data Analysis*, Wiley, New York.
14. Parmar, A.B. and Machin, A. (1995). *Survival Analysis. A Practical Approach*. Wiley, New York.
15. Pedersen, T. M., (1999). The Hodrick-Prescott Filter, <http://www.econ.ku.dk/tmp/manuscript.htm>
16. Rhodes, J. and Khan, A. (1971). *Office Dispersal and Regional Policy*, Cambridge University Press.
17. Rostow, W.W. (1960). *The Stages of Economic Growth*. Cambridge University Press, London.
18. Thom, R. (1975). *Structural Stability and Morphogenesis*, Benjamin, New York.
19. Townroe, P.M. (1971). *Industrial Location Decisions. A Study in Management Behavior*, Center for Urban and Regional Studies, The University of Birmingham.
20. Virras, I. (1999). *Regional Development Issues. A Survival Analysis Approach*. Unpublished M.Sc. Thesis, Department of Statistics, Athens University of Economics and Business.
21. Zeeman (1973). Applications of Catastrophe Theory. *Proceedings Volume, International Conference on Manifolds*. Tokyo.