



**UNIVERSITI PUTRA MALAYSIA**

**SURVIVAL ANALYSIS OF FOOD SECURITY IN ASIAN COUNTRIES**

**ANWAR FITRIANTO**

**FPSK(M) 2005 18**

**SURVIVAL ANALYSIS OF  
FOOD SECURITY IN ASIAN COUNTRIES**

By

**ANWAR FITRIANTO**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,  
in Fulfilment of the Requirements for the Degree of Master of Science**

**April 2005**



## **DEDICATION**

This thesis is dedicated to my wife, Greiche Dian Kusumawardhani, and my sweet daughter, Khazbiika Shahrinaz Anwar. Someones who are not always being forgotten, my parent (Maksum and Sunifah), parent-in-law (Roelche Chairul Syahfri and Hermien Sulianthy), who have always believed in me, and my brother-in-law (Syahfreal Dion Kusumawardhana), my three elder sisters (Mutmainnah, Sulismiati, Sri Kusrini) and my two elder brothers (Imam Hanafi and Nurahmad Fauzi).

Almighty Allah blessed me a livelihood and grace through SEARCA for the scholarship.



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment  
of the requirement for degree of Master of Science

## **SURVIVAL ANALYSIS OF FOOD SECURITY IN ASIAN COUNTRIES**

By

**ANWAR FITRIANTO**

**April 2005**

**Chairman : Associate Professor Isa Bin Daud, PhD**

**Faculty : Science**

This study focuses on using the survival analysis on food security application. The technique examines the effects of covariates on food insecurity among Asian countries in the period of 40 years since 1961. The analysis is carried out in order to determine the 'warning sign' of food insecurity condition. The data sources are from FAO and World Bank online database which include some particulars of 32 Asian countries.

It is observed that 21 of 32 (65.62%) countries experienced insecurity food condition. The remaining are censored observations (34.38%). The stepwise Cox's regression is used to select among the 24 independent covariates that are deemed to be significant contribution to the model. Initial run of the SAS code finds that six covariates are significant.

Based on the adopted model, at each time point, the West Asian region are found to be more likely to have insecurity food condition compared to those countries in the other regions. Furthermore, the occurrence of food security for East Asia countries

are more likely than for those in the other region. Meanwhile, it can also be seen that countries in Lower-middle income group are more likely to reach insecurity food condition than those in the other group. The analysis also shows that the high income countries have high risk of exposure to insecurity food condition.

Since Cox regression analysis has the basic assumption of proportionality, the model was tested whether it meets this condition. We use graphical method and formal test of this assumption . In the presence of ties, the ties-handling method of Breslow, Efron, Exact, and Discrete are compared with respect to Wald statistics, parameter estimate, the hazard ratio, and p-value.

The availability of the determined dataset as in allows assessing categories of food insecurity; Low, Medium, or High, which is useful to describe the nature of the food insecurity conditions. Based on the analysis, we are able to find variables that play important role on each stage of food insecurity condition of each country.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia  
sebagai memenuhi keperluan untuk ijazah Master Sains

## **ANALISIS MANDIRIAN ATAS KESELAMATAN MAKANAN NEGARA-NEGARA ASIA**

Oleh

**ANWAR FITRIANTO**

**April 2005**

**Pengerusi : Profesor Madya Isa Bin Daud, PhD**

**Fakulti : Sains**

Kajian ini tertumpu kepada penggunaan analisis mandirian bagi aplikasi keselamatan makanan. Teknik ini memeriksa pengaruh kovariat bagi ketakselamatan makanan di antara negara-negara Asia selama 40 tahun sejak tahun 1961. Analisis dilaksanakan bagi menentukan tanda amaran akan keadaan ketakselamatan. Data bersumber daripada FAO dan pangkalan data dalam talian Bank Dunia yang meliputi beberapa pembolehubah dari 32 negara-negara Asia.

Daripada 32 negara, 22 negara (65.72%) mengalami keadaan takselamat makanan. Selebihnya merupakan data tertapis (34.38%). Regresi Cox bertingkat digunakan untuk memilih kovariat yang dianggap sebagai penyumbang kepada model dari kesemua 24 kovariat. Berdasarkan operasi yang dijalankan dengan program SAS, didapati 6 kovariat adalah signifikan.

Berasaskan pada model yang digunapakai, pada setiap titik masa, rantau Asia Barat lebih cenderung mengalami keadaan ketidakselamatan makanan berbanding dengan

negara-negara di rantau lain. Sebaliknya, kejadian keselamatan makanan bagi rantau Asia Timur mempunyai kebarangkalian lebih besar berbanding dengan rantau lain. Sementara itu, dapat juga diperhatikan bahawa negara dengan pendapatan Sederhana-rendah mempunyai kebarangkalian lebih besar dalam mencapai keadaan ketakselamatan makanan berbanding dengan hal serupa bagi negara berpendapatan lain. Analisis juga menunjukkan bahawa negara berpendapatan tinggi mempunyai risiko lebih tinggi untuk terdedah kepada keadaan ketidakselamatan makanan.

Oleh kerana analisis Regresi Cox memiliki anggapan asas kekadaran, maka model diujikaji apakah memenuhi syarat kekadaran. Kami menggunakan kaedah grafik dan ujikaji rasmi untuk anggapan ini. Pada kehadiran seri, kaedah mengawal seri dari Breslow, Efron, Exact, or Diskrit dibandingkan dalam hal statistik Wald, anggaran parameter, nisbah bahaya, dan nilai-p.

Set data sedia ada dan kovariat memungkinkan penaksiran kategori bagi tahap keselamatan makanan samada Rendah, Sederhana, dan Tinggi, yang berguna untuk menjelaskan sifat keadaan ketakselamatan makanan. Berdasarkan kepada analisis berkenaan, kami dapat menemukan pembolehubah yang berperanan penting pada setiap tahap keadaan ketakselamatan makanan bagi setiap negara.

## ACKNOWLEDGEMENTS

In the name of Allah, the most gracious and merciful.

I would like to thank my thesis advisors Assoc. Prof. Isa Bin Daud, Ph.D, Assoc. Prof. Noor Akma Ibrahim, Ph.D, and Assoc. Prof. Mohd Rizam Abu Bakar, Ph.D who underwent with me through many of the labor intense moments in the delivery of this work.

This research was supported by the South East Asia Ministry of Education Organization – South East Asia Research for Agricultural (SEAMEO-SEARCA) of program award year 2003-2005, which is funded by German Academic Exchange Services (DAAD).



## TABLE OF CONTENTS

	Page
<b>DEDICATION</b>	ii
<b>ABSTRACT</b>	iii
<b>ABSTRAK</b>	v
<b>ACKNOWLEDGEMENTS</b>	vii
<b>APPROVAL</b>	viii
<b>DECLARATION</b>	x
<b>TABLE OF CONTENTS</b>	xi
<b>LIST OF TABLES</b>	xiv
<b>LIST OF FIGURES</b>	xv

## CHAPTERS

<b>I</b>	<b>INTRODUCTION AND OVERVIEW</b>	
1.1	Background	1
	1.1.1 Asia Countries' Population Growth and Food Pressure	1
1.2	Objectives	3
<b>II</b>	<b>LITERATURE REVIEW</b>	
2.2	Survival Analysis	4
	2.1.1 The History of Survival Analysis	6
	2.1.2 Survival Analysis Applications	7
2.2	Some Tools Used in Survival Analysis	8
	2.2.1 The Cumulative Distribution Function	8
	2.2.2 The Probability Density Function	9
	2.2.3 The Survival Function	9
	2.2.4 The Hazard Function	10
	2.2.5 Cox's Proportional Hazards Model	12
	2.2.6 The Assumptions of Proportional Hazard	16
	2.2.7 The Treatment of Ties	17
	2.2.8 Residual Analysis and Diagnostics	19
2.3	Origins of Censoring	21
	2.3.1 Types of Censoring Mechanism	21
	2.3.1.1 Type I Censoring (Time Censoring)	22
	2.3.1.2 Type II Censoring	23
	2.3.1.3 Type III Censoring (Random Censoring)	23
	2.3.2 Common Statistical Methods for Censored Data	25



	2.3.2.1	Complete-Data Analysis	25
	2.3.2.2	Imputation Approach	27
	2.3.2.3	Analysis Based on Dichotomized Data	28
	2.3.2.4	Likelihood-Based Approach	29
	2.3.3	Necessity of Making Assumptions about Censoring	29
<b>III</b>		<b>POPULATION AND FOOD SECURITY</b>	
	3.1	Population Changes	34
	3.2	Food Security	35
	3.2.1	Food Demand Issues	36
	3.2.2	Food Supply Issues	37
	3.3	Natural Resource Connections to Food Security	38
	3.3.1	The Environment and Food Security	40
	3.3.2	Forest and Food Security	40
	3.3.3	Fertilizer and Food Security	41
	3.3.4	Water and Food Security	41
	3.3.5	Women and Food Security	42
<b>IV</b>		<b>DATA AND METHODOLOGY</b>	
	4.1	Research Questions	43
	4.2	Data Description	43
	4.2.1	Variables Definition	47
	4.2.2	Dummy Variables	47
	4.3	Methods	48
	4.3.1	Creating Time Variable and Event Definition	49
	4.4	Using SAS	52
<b>V</b>		<b>RESULTS</b>	
	5.1	Data Exploration	54
	5.1.1	Agricultural Production Indices	54
	5.1.2	Region	56
	5.1.3	Human Resources	57
	5.1.4	Irrigation to Agricultural Area	57
	5.1.5	Land use for Agricultural and Total Area	59
	5.2	Cox's Model Building	61
	5.2.1	Cause-Specific Survival Probability	67
	5.2.2	Assessment of Food Insecurity Categories	69
	5.3	Test of The Proportionality Assumption and Residual Diagnostics	71
	5.3.1	Graphical Method of Proportionality	71
	5.3.2	Formal Test of Proportionality	72
	5.3.3	Residual Diagnostics	74
	5.4	Effects of Ties-Handling Methods	76

<b>VI</b>	<b>CONCLUSION AND DISCUSSION</b>	
6.1	Conclusion	78
6.2	Discussion and Extension	80
	<b>REFERENCES</b>	83
	<b>APPENDICES</b>	87
	<b>BIODATA OF THE AUTHOR</b>	121

## LIST OF TABLES

Table		Page
1	Several application fields of survival analysis (Smith, 2003)	7
2	List of variables included in the analysis	46
3	Categorical variable and the coding	48
4	Illustration of event and time definition based on WAPI data	51
5	Geographically distribution of Asian countries under studies	56
6	Chi-square tests of Asian countries in association with region	56
7	Parameter estimation of the Cox's model of Asia countries food security	63
8	Cross tabulation between region of Asian countries and stage of food insecurity	69
9	Chi square tests of stage of food insecurity Asian countries in association with region	70
10	Categorical variable building of the Asian countries food insecurity stages	71
11	Testing the proportional hazards assumption using a time dependent covariate	73
E.1	Summary of censoring observation at 1%-45% reduction of weighted Agricultural Production Indices	108

## LIST OF FIGURES

Figure		Page
1	Survival analysis : main distributional representation	11
2	Sustainable agriculture and food security: making the link	39
3	Trend of weighted Agricultural Production Indices of some selected Asia countries under study during 1961-2001	50
4	Illustration of study design where the observation times start at a consistent point in time ( $t = 0$ )	52
5	Time series data of Agricultural Production Indices of Asia countries during 1961-2001	55
6	Time series data of weighted Agricultural Production Indices of Asia countries during 1961-2001	55
7	Distribution of human development categories among regions in the world	57
8	Trend of irrigation to agricultural area among Asian countries during 1961-2001	58
9	Trend of tractor usage among Asian countries during 1961-2001	59
10	Percentages trend of land usage which is used to agricultural purposes of Asian countries during 1961-2001	60
11	Trend of land usage which is used to arable and permanent crops of Asian countries during 1961-2001	60
12	Comparison for West Asia and the other countries of the estimated hazard function $\hat{\lambda}(t)$ of time to insecurity food condition	65
13	Comparison for East Asia and the other countries of the estimated hazard function $\hat{\lambda}(t)$ of time to insecurity food condition	66
14	Comparison for the High income and the other countries of the estimated hazard function $\hat{\lambda}(t)$ of time to insecurity food condition	66
15	Thirty years estimated cause-specific survival probability curves of Asian countries of some covariates, obtained by using a survival prediction method.	68
16	Plot of $\text{Log}[-\text{Log}(S(t))]$ against time	72
17	Martingale residual plot observations where event occurred	75
18	Deviance residual plot of observations where event occurred	75
D.1	Comparison of ties handling methods for Wald statistics	106

D.2	Comparison of ties handling methods for absolute value of the parameter estimate	106
D.3	Comparison of ties handling methods for p value of the parameter estimate	107
D.4	Comparison of ties handling methods for the hazard ratio	107

## **CHAPTER I**

### **INTRODUCTION AND OVERVIEW**

#### **1.1 Background**

The challenge of sustainable of food production in the 21st Century can be summarized through the following question: Can food productions keep in pace with the population growth, especially in parts of the world where population continue to grow rapidly?

##### **1.1.1 Asia Countries' Population Growth and Food Pressure**

In population terms, Asia contributes the most to world population growth, at 50 million people a year, while Africa accounts for only 17 million; although at 2.36 percent, Africa's rate of growth is the highest. Two of every five people alive today are living in China or India. While 10 nations currently have populations that exceed 100 million, the number of nation is expected to rise to 19 by 2050. Half out of these 10 countries are Asian countries (United Nation Population Fund, 2003). The variance in rates of population growth among individual countries will be responsible for substantial change in the top 10 contributors to world population growth in this century.

The United Nations observed that October 12, 1999 as the day of the Sixth Billion - the world's population had doubled since 1960. In some parts of the developing world the population grew even faster, e.g. in Sub-Saharan Africa it increased



threefold. Asia's number of people grew fastest in absolute terms: by nearly two billion. Based on IFPRI projection on 1995-2020 (Pinstrup-Andersen, et al., 1999), world population will increase by 32% to 7.5 billion, mostly in cities in developing countries. And 85% of total food demand growth will come from developing countries

The Food and Agriculture Organization of the United Nations (FAO) defines "food security" as a state of affairs where all people at all times have access to safe and nutritious food to maintain a healthy and productive life. Meanwhile, food is an essential requirement for every individual. Besides nourishing the biological needs, it helps to guarantee the welfare of the individual, serves to improve the productivity of the labor force, and hence reduce social expenditure, safeguard political stability. It is thus essential and of primary importance to ensure minimum levels of food security for the poor. Food insecurity is the result of a discrepancy between agricultural production and population growth. The major thrust of food security is to bring about a significant increase in agricultural production in a sustainable way and to achieve a substantial improvement in people's entitlement to adequate food and culturally appropriate food supplies. If the condition is not seriously observed, it is not impossible that world hunger will occur. In other words, the agricultural sustainability must be kept in order to obtain food security.

The analysis of survival data as proposed in this research is the basis for a perspective on world food security risk assessment. The objective is to contribute to the understanding of the dynamic processes that underlies the agricultural food



production in Asian face as a result of rapid population growth. The method is to estimate the ‘survival function’ and use the developed model for duration data until ‘warning’ of food insecurity comes.

## **1.2 Objectives**

The objectives of this thesis are:

- to provide alternative solution on food problem of Asian countries through statistical analysis.,
- to adopt a mathematical model that describes the survivorship of Asian countries on food security,
- to find out factors which influence difference of hazard ratio among Asian countries on food security, so that the insecurity food condition can be detected early,
- to classify current level of food insecurity condition of Asian countries, and to identify factors which influence the condition at each level.
- to predict the survival pattern of different characteristics of the Asian countries.

## **CHAPTER II**

### **LITERATURE REVIEW**

#### **2.1 Survival Analysis**

A class of statistical analysis methods, termed “survival analysis,” is increasingly used with great frequency in the medical literature. It consists of methods of studying occurrence and timing of events (Lee, 1992). This leads us to the question of what is an event. An event is a qualitative change that is of interest occurring at some points on the time line.

The goal of survival analysis is to analyse measures describing in some sense the width of the interval between an origin point and an end point. Often, the end point corresponds to death or culling and the length from the origin to the end is measured in the time units (seconds, minutes, days, months or years).

A survival analysis problem needs:

1. a starting point of event. This needs to be well-defined.
2. a scale or method for measuring time. This could be conventional methods such as minutes, days, or years. In some applications a different concept of time may be more applicable. When purchasing a used car the mileage the car is driven is often more relevant to the operability of a car than the calendar time since the car was manufactured.
3. a stopping time point. Again, this needs to be well-defined.

A common example is the life span of human life. The date of birth is often the beginning event, calendar time is the usual time scale, and death is a common ending event. In statistical terms we are interested in the random variable  $T$  defined as the time elapsed from the starting event to the occurrence ending event. We are usually interested in three kinds of questions involving this random variable; how to characterize the random variable, how to compare survival among groups, and finally how to determine factors or study predictor variables related to survival.

The use of survival analysis, as opposed to the use of a different statistical method, is most important when some subjects are lost to follow-up or when the period of observation is finite, such that not all patients experience the event of interest during the study period. For this case, the actual time measurements are not really observed. However, if the distribution of the variable is known, we may resort the probability of the actual being greater than the observed. These cases constitute the censored observation. In many clinical trials, not all individuals recruited in the study are completed up to the end of the study. Thus, we have the time variable being censored when the study is incomplete and uncensored if otherwise.

Survival analysis methods are useful, however, for analyzing any time-to-event data. The event of interest, traditionally death, can be replaced with any endpoint that occurs at a particular time, and can occur only once. It is the duration of time until the endpoint that is of interest. The end point (generally called failure) may also correspond to the occurrence of any type of event such as ringgit or dollars spent or earned, kilograms of milk produced, litters born, etc (Durocq, 1997).

Some of the common applications now include time until onset of disease, time until stock market crash, time until equipment failure, time until earthquake, and so on. The best way to define such events is simply to realize that these events are a transition from one discrete state to another at an instantaneous moment in time.

### **2.1.1 The History of Survival Analysis**

The origin of survival analysis goes back to mortality tables from centuries ago. However, it was not until World War II that a new era of survival analysis emerged. This new era was stimulated by interest in reliability (or failure time) of military equipment. At the end of the war these newly developed statistical methods emerging from strict mortality data research to failure time research, quickly spread through private industry as customers became more demanding of safer, more reliable products. As the uses of survival analysis grew, parametric models gave way to nonparametric and semi parametric approaches for their appeal in dealing with the ever-growing field of clinical trials in medical research. Survival analysis was well suited for such work because medical intervention follow-up studies could start without all experimental units enrolled at start of observation time and could end before all experimental units had experienced an event. This is extremely important because even in the best-developed studies, there will be subjects who choose to quit participating, who move too far away to follow, or who will die from some unrelated event.

The researcher was no longer forced to withdraw the experimental unit and all its associate covariates from the study instead techniques called censoring and enabled

researchers to analyze incomplete data due to delayed entry or withdrawal from the study. This was important in allowing each experimental unit to contribute all of the information possible to the model for the amount of time the researcher was able to observe the unit. The last great strides in the application of survival analysis techniques has been a direct result of the availability of software packages and high performance computers which are now able to run these difficult and computationally intensive algorithms relatively efficient.

### 2.1.2 Survival Analysis Applications

Smith (2003) mentioned several field that survival analysis may be applicable. In the following table are some applications of his suggestion.

Table 1: Several application fields of survival analysis (Smith, 2003)

No.	Field of Application
1	<u>Medical Field</u> <ul style="list-style-type: none"> <li>• Death</li> <li>• Relapse</li> <li>• Occurrence of symptoms</li> <li>• Disease onset</li> </ul>
2	<u>Reliability</u> <ul style="list-style-type: none"> <li>• Product failure</li> <li>• Machine repair</li> </ul>
3	<u>Sociology</u> <ul style="list-style-type: none"> <li>• Divorce</li> <li>• Career change</li> <li>• Smoking cessation</li> <li>• First marijuana use</li> </ul>
4	<u>Business/Economics</u> <ul style="list-style-type: none"> <li>• Bankruptcy</li> <li>• Unemployment assistance</li> <li>• Divestiture of stocks</li> <li>• Labor strike duration</li> </ul>

Of course, in the real world, there is no limitation in using survival analysis methods to solve problems other than the applications suggested by Smith.

## 2.2 Some Tools Used in Survival Analysis

Let  $T$  be a nonnegative random variable representing the failure time of an individual from a population. The distribution of  $T$  can be specified in many ways, three of which are particularly useful in survival applications: the probability density function, the survivor function, and the hazard function (Lee, 1992; Deshpande and Sudha, 2001). Although these three functions are mathematically equivalent, if one of them is given, the other two can be derived. In practice, the three survival functions can be used to illustrate different aspect of the data.

### 2.2.1 The Cumulative Distribution Function

The cumulative distribution function (cdf) is very useful in describing the continuous probability distribution of a random variable, such as time, in a survival analysis. The cdf of a random variable  $T$ , denoted  $F_T(t)$ , is defined by

$$F_T(t) = P(T < t); \quad t \geq 0. \quad (2.1)$$

This is interpreted as a function that will give the probability that the random variable  $T$  will be less than to any value  $t$  that we choose. Several properties of a distribution function  $F_T(t)$  can be listed as a consequence of the knowledge of probabilities. Because  $F_T(t)$  has the probability  $0 \leq F_T(t) \leq 1$ , then  $F_T(t)$  is a non decreasing function of  $t$ , and as  $t$  approaches  $\infty$  then  $F_T(t)$  approaches 1.