Cox Regression Models with Time-Varying Covariates Applied to Survival Success of Young Firms (*)

Aygül Anavatan, Murat Karaöz

Akdeniz University, İİBF, Department of Econometrics

07058, Kampus, Antalya, Turkey

E-mails: *aygulanavatan@akdeniz.edu.tr*, *mkaraoz@akdeniz.edu.tr*

Abstract

Cox proportional hazards model assumes that independent variables remain constant throughout the observation period. Model can give biased results in cases which this assumption is violated. One of the methods used modelling the hazard ratio in the cases that the proportional hazard assumption is not met is to add a time-dependent variable showing the interaction between the predictor variable as parametric function of time. In this study, we investigate the factors that affect the survival time of the firms and the time dependence of these factors using Cox regression considering time depedent independent variables.

(*) This paper is an extension to the findings of the scientific research project "The Factors Affecting Survival and Growth Performance of Newly Established Enterprises in Business Incubators: A Survey on the KOSGEB Business Development Centers (İŞGEM)", 109K139, which has been funded with grant from TÜBİTAK. We also acknowledge the administrative support to the project from KOSGEB.

Keywords: Survival analysis; Cox Regression Model; Proportional Hazard Assumption; New Firms

1.INTRODUCTION

Survival analysis deals with the probability of occurrence of a given event at a set of particular points in a time interval (Cox and Oakes, 1984). The typical survival anaylsis may include the reports of hazard rates, ratios and survival curves while relating a likely set of independent variables to a specific event. In the small business and entrepreneurship literature, survival analysis has been used to track the start-ups over the years. A survival curve of a cohort of newly established firms reports what percentage of the cohort continue to survive since its inception over time, indicating whether some of the firms are failed over the years (Karaöz and Albeni, 2011). Cox proportional hazards (PH) model is the most preferred model in order to investigate the effect of variables on survival time. The key assumption of Cox model is that hazard rate related to different levels of the factors is constant throughout the follow-up period (Başar, 2006). Violation of the PH assumption requires additional measures for unbiased results of Cox Survival regression. In this paper, Cox regression has been applied to investigate the survival of newly established firms under incubation. Violation of PH assumption has been tested and further Cox regressions are performed considering time-varying effects of independent variables to survival.

2.SURVIVAL ANALYSIS

In a survival analysis, it is usually referred to the time variable as survival time, because it gives the time that an individual has "survived" over some followup period (Geiss et al., 2009). It is also typically referred to the event as a failure, because the event of interest usually is death, disease incidence, or some other negative individual experience (Kleinbaum and Klein, 2005).

When survival time (T) is defined as a random variable with cumulative distribution function $\mathrm{P(t) = Pr(T \leq t)}$ and probability density function $\mathrm{P(t) = d\,P(t)/d\,(t)}$, survival function $\mathrm{S(t)}$ is explained by Equation (2.1) (Yay, Çoker and Uysal, 2007);

$$\mathrm{S(t)} = \mathrm{P}(\mathrm{T} > t) = 1 - \mathrm{P(t)} \qquad (2.1)$$

Survival function $\mathrm{S(t)}$ gives the probability that the random variable T exceeds the specified time t (Kleinbaum and Klein, 2005). All survival functions have the characteristics that i) they are nonincreasing; that is, they head downward as t increases, ii) at time $\mathrm{t} = 0$, $\mathrm{S(t) = S(0)} = 1$; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one, iii) at time $\mathrm{t} = \infty$, $\mathrm{S(t) = S(\infty)} = 0$; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survival curve must eventually fall to zero (Kleinbaum and Klein, 2005).

The hazard function $\mathrm{h(t)}$, with its complement of survival function $\mathrm{S(t)}$, is given by Equation (2.2), where $\Delta\mathrm{t}$ denotes a small interval of time (Kleinbaum and Klein, 2005);

$$\mathrm{h(t)} = \lim_{\Delta \mathrm{t} \to 0} \frac{\mathrm{P}(\mathrm{t} \leq \mathrm{T} < t + \Delta t \mid \mathrm{T} \geq \mathrm{t})}{\Delta \mathrm{t}}$$

(2.2)

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 (Tabatabai et al., 2007). In contrast to the survival function, which focuses on not failing, the hazard function focuses on failing, that is, on the event occurring (Kleinbaum and Klein, 2005).

2.1.The Cox Proportional Hazards (PH) Model

The Cox PH model is usually written in terms of the hazard model formula shown at equation (2.3). This model gives an expression for the hazard at time t for an individual with a given specification of a set of explanatory variables denoted by X. That is, X represents a collection of predictor variables that is being modeled to predict an individual's hazard (Kleinbaum and Klein, 2005).

$$h(t, X) = h_0(t)e^{\sum_{i=1}^{p} \beta_i X_i}$$

(2.3)

The Cox model formula says that the hazard at time t is the product of two quantities. The first of these, $h_0(t)$, is called the baseline hazard function. The second quantity is the exponential expression e to the linear sum of $\beta_i X_i$, where the sum is over the p explanatory X variables (Kleinbaum and Klein, 2005). A hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. The two individuals being compared can be distinguished by their values for the set of predictors, that is, the X's. HR is shown by the following formula, where X^* denotes the set of predictors for one individual, and X denotes the set of predictors for the other individual (Kleinbaum and Klein, 2005);

$$\widehat{HR} = \frac{\hat{h}(t, X^*)}{\hat{h}(t, X)} = \frac{\hat{h}_0(t) \exp[\sum \hat{\beta}_i X_i^*]}{\hat{h}_0(t) \exp[\sum \hat{\beta}_i X_i]} = \exp\left[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)\right]$$

(2.4)

Once the model is fitted and the values for X^* and X are specified, the value of the exponential expression for the estimated HR is a constant, $\hat{\theta}$, which does not depend on time (Kleinbaum and Klein, 2005);

$$\hat{\theta} = \exp\left[\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)\right]$$

(2.5)

Running the Cox regression, observations should be independent of each other and HR should remains constant with time. This assumption related to hazard ratio is known as PH assumption. If the HR is increasing over time, the estimated coefficients assuming PH is overestimating at first and underestimating later on (Bellera et al., 2010).

2.2.Extension of the Cox Proportional Hazards Model

An important feature of this formula, which concerns the PH assumption, is that the baseline hazard is a function of t, but does not involve the X's. The X's in the formula are called time-independent X's (Kleinbaum and Klein, 2005). It is possible, nevertheless, to consider X's which do involve t. Such X's are called time-dependent variables. If time-dependent variables are considered, the Cox model form may still be used yet in an extended form, as the orginal model do not satify the PH assumption (Kleinbaum and Klein, 2005).

In the case of being time-dependent explanatory variables, Cox regression model expands to a model which contains time-independent variables and some functions of the time the product with these variables. Independent variables are, $X(t)$, where $X_1, X_2, \ldots, X_{p_1}$ time-independent variables and $X_1(t), X_2(t), \ldots, X_{p_2}(t)$ time-dependent variables (Sertkaya et al., 2005). Then, Cox regression model is, β and δ which denote vector of coefficients of explanatory variables (Sertkaya et al., 2005);

$$h(t, X(t)) = h_0(t) \exp\left[\sum_{i=1}^{p_1} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_j g(t)\right] \quad (2.6)$$

Where $g(t)$ is defined as a function of time. $g(t)$ usually is defined in the form of t, $\log(t)$, $\ln(t)$ or as step function (Sertkaya et al., 2005).

3.AN APPLICATION INTO NEW FIRM SURVIVAL UNDER INCUBATION

Although the Survival analysis extensively been used in medical research on individuals, recently it becomes widely popular in Business Success and survival research. Thus, rather than on individuals, in this paper, we apply Cox regression to investigate the survival of newly established firms under incubation. There are studies applying survival Violation of PH assumption has been tested and further Cox regressions are performed considering time-varying effects of independent variables to survival. Our 414 observations on firm characteristics acquired from 12 different incubators, İŞGEMs, located across Turkey. The data includes almost all firms that currently exist İŞGEMs or the firms that resided in the past yet left İŞGEMs by graduation or failure. The survey data consists the total of.

A business incubator can be identified as an organization which mentors the development of newly founded firms by specialized services such as providing office space, specialized staff, machinery, equipment, facilities and business assistance (Aernoudt, 2004).

3.1.Variables Used in the Analysis

For our analysis, factors affecting the initial success of young enterprises can be summarized as i) Human capital characteristics of new enterprise's owner such as education level and sector experience, ii) Firm characteristics such as scale, age and human capital, iii) Industry characteristics such as market growth rate and entry barriers, vi) Incubation features, v) Other external factors such as macroeconomic fluctuations, regional factors and public policies (Hackett and Dilts, 2004). All of the data and variables used in our analysis are taken from Karaöz and Albeni (2011) and descriptive statistics and definitions are presented at Table 3.1.

Table 3.1: The variables used in analysis and descriptive statistics

VARIABLE	DEFINITION	Observations	Mean	Minimum	Maximum
	EVENT OF INTEREST				
exit	If the firm is closed (failed) during or after the incubation 1, otherwise 0	414	-	0	1

		DEPENDENT VARIABLE				
incubage		The elapsed time from the firm's entry into incubation until it's closed (month)	404	41,52	2	158
		INDEPENDENT VARIABLES				
CHARACTERISTICS OF THE ENTREPRENEUR	**income**	If entrepreneur's income only comes from the incubated firm 1, otherwise 0	414	-	0	1
	gender	If entrepreneur is female 1, male 0 (If there are both male and female partner 0)	414	-	0	1
	lnentage	Entrepreneur's age (If there is a partnership, it is taken as the oldest entrepreneur's age-logarithmic scale)	367	3,64	3	4,25
	enteduuni	If entrepreneur is a college graduate 1, otherwise 0 (if there is a partnership and one of the partners is college graduate 1)	414	-	0	1
	entexp	Entrepreneur's prior experience before arriving İŞGEM (year)	414	5,83	0	40
	family	If there is a role model for entrepreneurship in entrepreneur's family or surrounding 1, otherwise 0	414	-	0	1
FEATURES OF THE FIRM	**partner**	The number of partners witin the established firm	414	1,24	1	4
	export	If the firm export 1, otherwise 0	414	-	0	1
	lnempini	initial firm size (logarithmic scale)	392	1,31	0	5,70
	onlyloan	If firm's founding capital is completely loan 1, otherwise	414	-	0	1

		0				
	networking	If entrepreneur is in cooperation with stakeholders within and outside the incubator 1, otherwise 0	414	-	0	1
	innova	If entrepreneur has made innovation 1, otherwise 0	414	-	0	1
	advert	If the firm has had an advertising 1, otherwise 0	414	-	0	1
	brand	If the firm is a brand owner 1, otherwise 0	414	-	0	1
INCUBATION SERVICES AND PROPERTIES	**comserv**	If entrepreneur has used at least one of the common services offered by incubation 1, otherwise 0	414	-	0	1
	whenest	If the incubated firm entered the incubation center within first 3 years (36 months) of incubation center 1, otherwise 0	414	-	0	1
	incubsize	The number of incubation's workshop	414	43,14	14	84
INDUSTRIAL PROPERTIES	**sector**	If the firm is in the manufacturing industry 1, in the service sector 0	411	-	0	1
	compete	Intensity of competition in the sector (1-5 Likert scale)	410	-	1	5
EXTERNAL FEATURES	**prorank**	(%) Share of the GDP per capita of the province in the Country GDP where the incubation center is located	414	1,51	0,59	2,07
	cycle	If the firm has experienced an economic crisis 1, otherwise 0	414	-	0	1

Source: Karaöz and Albeni (2011).

According to our data, Firm Survival Curve has been presented at Figure 3; Surivors diminish to about 20% with 158 months.

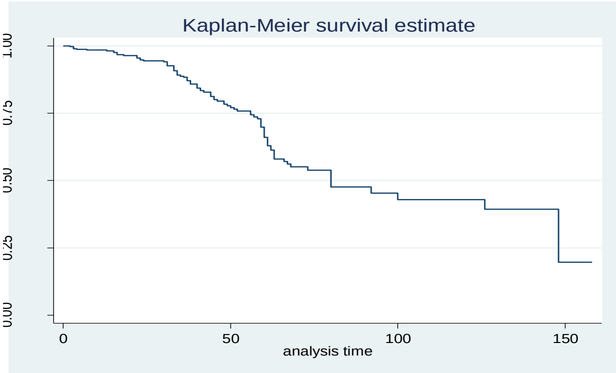

Figure 3.1: Survival curve of firms with failures during or after incubation

3.2. Results

All Cox Regression results with and without considering time effects are presented in Table 3.2. Our PH tests indicate that further estimations are necessary using time-dependent variables. Model 2 estimates include the variables which in Model 1 and all of the interaction terms created by each of these variables multiplying $\ln(t)$, which is a function of time, in order to handle variable-time interaction. The Model 3 are obtained by using only the relevant variables of (incubsize) and (prorank), which are found to be the time-dependent variables.

Table 3.2: Estimates of the basic and with time-dependent Cox model variables

	Coefficients $P > \|z\|$																				
Variable	**income**	**gender**	**lnentage**	**enteduuni**	**entexp**	**family**	**partner**	**export**	**lnempini**	**onlyloan**	**networking**	**innova**	**advert**	**brand**	**comserv**	**whenest**	**incubsize**	**sector**	**compete**	**prorank**	**cycle**
Model 1	1.18	-0.056	0.265	0.659	-0.084	-0.307	-1.71	0.827	0.214	-1.03	-1.47	-1.67	0.636	0.865	0.264	-1.18	-0.02	-0.156	-0.157	-1.16	0.46
	0.010***	0.892	0.732	0.044**	0.042**	0.402	0.015**	0.308	0.278	0.063*	0.004***	0.006***	0.17	0.275	0.592	0.013**	0.002***	0.738	0.416	0.013**	0.21
Model 2	5.16	1.14	6.45	0.289	-0.307	-4.63	1.07	6.47	0.274	-6.97	4.45	-2.24	-2.74	-4.22	-6.63	-0.985	-0.261	-4.39	2.92	19.4	10.5
	0.253	0.745	0.422	0.924	0.342	0.205	0.844	0.576	0.847	0.215	0.333	0.74	0.522	0.685	0.217	0.813	0.015**	0.291	0.23	0.003***	0.085*
Model 3	1.75	-0.093	0.721	0.762	-0.101	-0.249	-2.39	0.951	0.196	-1.88	-1.54	-2.46	0.615	1.61	0.638	-2.25	-0.253	-0.425	-0.341	16.9	0.791
	0.000***	0.826	0.364	0.024**	0.013**	0.518	0.001***	0.303	0.298	0.002***	0.004***	0.001***	0.198	0.074*	0.234	0.000***	0.009***	0.362	0.099*	0.002***	0.040**
	income $\times$ **ln**(t)	**gender** $\times$ **ln**(t)	**lnentage** $\times$ **ln**(t)	**enteduuni** $\times$ **ln**(t)	**entexp** $\times$ **ln**(t)	**family** $\times$ **ln**(t)	**partner** $\times$ **ln**(t)	**export** $\times$ **ln**(t)	**lnempini** $\times$ **ln**(t)	**onlyloan** $\times$ **ln**(t)	**networking** $\times$ **ln**(t)	**innova** $\times$ **ln**(t)	**advert** $\times$ **ln**(t)	**brand** $\times$ **ln**(t)	**comserv** $\times$ **ln**(t)	**whenest** $\times$ **ln**(t)	**incubsize** $\times$ **ln**(t)	**sector** $\times$ **ln**(t)	**compete** $\times$ **ln**(t)	**prorank** $\times$ **ln**(t)	**cycle** $\times$ **ln**(t)
Model 2 (cont.)	-0.959	-0.352	-1.48	0.139	0.055	1.23	-1.08	-1.25	-0.023	1.3	-1.7	-0.147	0.924	1.48	1.99	-0.415	0.059	1.03	-0.903	-5.74	-2.5
	0.427	0.699	0.474	0.861	0.505	0.212	0.457	0.678	0.954	0.378	0.177	0.937	0.418	0.582	0.161	0.718	0.030**	0.344	0.17	0.001***	0.108

Model 3 (cont.)																	0.058			-5	
																	0.017**			0.001***	

*, **, and *** indicate significance at the 1, 5 and 10% levels, respectively.

Log-likelihood and prob values of Model 1, 2 and 3, respectively, are -190.632 [0.000***], -165.552 [0.000***] and -173.255 [0.000***].

Considering all the model estimates, we obtain various results regarding the variables. The possibility of failure of the firms, whose owners only dependent on earnings coming from its new-born firm higher than other firms. In this case it has been seen that the entrepreneurs having income from other sources are more likely to be successful in start-up business. It is interesting to see the result that the firms whose owners are university graduates have about two times higher risk of failure than other firms. An increase in the number of partners in the firm decreases the possibility of failure of firms. It is interesting to see that failure risk of firms, whose founding capital is formed entirely by loans, are lower than whose initial capital is partially or fully self-financed. If an entrepreneur is in collaboration with stakeholders within and outside the incubation, survival probability of the firm becomes higher. Moreover, it has been seen from the estimates that innovation activity of new firms increases chance of survival. Brand ownership also increases the chance of the firm's survival. Establishing a firm within an incubation center that is within its first 3-years increases survival probability. Finally, firms those experience a macroeconomic crisis have nearly two times more likelyhood of failure than others.

4.CONCLUSIONS

Cox proportional hazard model, besides others, rest on proportional hazards assumption that independent variables do not vary with time. When PH assumption is violated and Cox regression estimates become biased. Then, Cox survival estimates can be corrected by including the time-varying effects to the analysis. Identification and calculation of time-dependent effects give the oppotunity to obtain some otherwise unseen valuable special time pattern information. In our analysis, initially, the Cox regression was performed by considering that all explanatory variables are constant over time. Then, extended Cox regression models were estimated by including the time-dependent explanatory variables in the model. Our extended model results have shown that it become usefull to estimate the Cox Proportional Hazards regression by also including the time-varying explanatory variables to the analysis. Both the time-independent and time-dependent variables create significant effects on the probability of survival of the İŞGEM firms.

REFERENCES

Aernoudt, R. (2004). Incubators: Tool for Entrepreneurship?. Small Business Economics, 23, 127-35.

Başar, E. (2006). Orantılı Olmayan Hazard Üzerine Bir Çalışma. Paper presented at 5. İstatistik Günleri Sempozyumu, Antalya,111-16.

Bellera, C.A., MacGrogan, G., Debled, M., De Lara, C.T., Brouste, V., & Mathoulin-Pélissier, S. (2010). Variables with time-varying effects and the Cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Medical Research Methodology, 10-20.

Cox, D.R., & Oakes, D. (1984). Anaylsis of Survival Data. London: Chapman and Hall.

Geiss, K., Meyer, M., Radespiel-Tröger, M., & Gefeller, O. (2009). SURVSOFT—Software for nonparametric survival analysis. Computer Methods and Programs in Biomedicine, Elsevier Ireland LTD., 96, 63–71.

Hackett, M., & Dilts, D.M. (2004). A Systematic Review of Business Incubation Research. Journal of Technology Transfer, 29, 55-82.

Kleinbaum, D.G., & Klein, M. (2005). Survival Analysis: A Self-Learning Text (2nd Ed.). New York: Springer.

Sertkaya, D., Ata, N., & Sözer, M. T. (2005). Yaşam çözümlemesinde zamana bağlı açıklayıcı değişkenli Cox regresyon modeli. Ankara Üniversitesi Tıp Fakültesi Mecmuası, 58, 153-58.

Tabatabai, M. A., Bursac, Z., Williams, D. K., & Singh, K. P. (2007). Hypertabastic survival model. Theoretical Biology and Medical Modelling, 4-40.

Karaöz, M., & Albeni, M., (2011), İş Kuluçkalarında Yeni Kurulan Girişimlerin Hayatta Kalma ve Büyüme Performansını Etkileyen Faktörler: KOSGEB İş Geliştirme Merkezleri (İŞGEM) Üzerine Bir Araştırma. The Scientific and Technological Research Council of Turkey (TÜBİTAK). (Issue Brief No. 109K139).

Yay, M., Çoker, E., & Uysal, Ö. (2007). Yaşam Analizinde Cox Regresyon Modeli and Artıkların İncelenmesi, Cerrahpaşa Tıp Dergisi, 38, 139-45.