ECONOMICS BULLETIN

Survival Analysis for Unobserved Heterogeneity on Estimated Mortality in Taiwan

Wen-Shai Hung
Department of Business Administration, Providence University, Taiwan

Shu-Hsi Ho
Department of International Business, Ling-Tung University, Taiwan

Abstract

The purpose of this study provides evidences on the effects of unobserved heterogeneity on estimated mortality among the middle aged and elderly in Taiwan. The data used is from the Survey of Health and Living Status of the Middle Aged and Elderly in Taiwan (aged 50 to 66), and the mortality information was linked to 1996-2003 national death registry data. The Weibull models are used to estimate the effects of unobserved heterogeneity on mortality. Main empirical results confirm that, after considering unobserved heterogeneity, most estimated coefficients on the mortality regressors are larger in magnitude that the corresponding coefficients in the reference model. Especially, the terms of health care utilization have larger unobserved heterogenity on estimated mortality. Therefore, if the government policies can concern more unobserved heterogenity of health care utilization that might useful to decrease the mortality for the elderly.

The authors would like to thank the Bureau of Health Promotion at the Department of Health for providing the Survey of Health and Living Status of the Middle Aged and Elderly in Taiwan data.
Citation: Hung, Wen-Shai and Shu-Hsi Ho, (2008) "Survival Analysis for Unobserved Heterogeneity on Estimated Mortality in Taiwan." *Economics Bulletin*, Vol. 9, No. 25 pp. 1-10
Submitted: July 24, 2008. **Accepted:** August 26, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume9/EB-08I00003A.pdf

Survival Analysis for Unobserved Heterogeneity on Estimated Mortality in Taiwan

1. Introduction

This paper investigates the factors of unobserved heterogeneity contributing to estimate the mortality among the middle aged and elderly in Taiwan. It is well known that survival analysis produces incorrect results if unobserved heterogeneity is ignored (Lancaster, 1990). However, previous studies of mortality issues in Taiwan by survival analysis only focused on the influencing factors of socio-demographic status, health self-assessment, physical functions, living arrangements, and health-care utilization; but never considered the effect of unobserved heterogeneity on estimated mortality (see, for example, Zimmer, Martin, and Lin, 2005; Lin and Lin, 2006; Ho, 2008). Therefore, this paper uses the Weibull model and considers unobserved heterogeneity to fill this gap (see, Cleves, Gould, and Gutierrez, 2002; Hosmer and Lemeshow, 1999).

First, this paper uses the data from the Survey of Health and Living Status of the Middle Aged and Elderly in Taiwan (SHLS), and the mortality information was linked to 1996-2003 national death registry data. Second, this study uses a cross-section estimation techniques to estimate the effects of unobserved heterogeneity on estimated mortality. The expected result can confirm that, after considering unobserved heterogeneity, most estimated coefficients on the mortality regressors are larger in magnitude that the corresponding coefficients in the reference model.

The remainder of the paper is structured as follows. Section 2 presents an overview of mortality estimation by survival analysis. Section 3 describes the estimation methods, including data source, variables specification, ethical considerations, and Weibull models without or with unobserved heterogeneity. This is followed by the major empirical results in Section 4, particularly for examining the effects of unobserved heterogeneity on mortality. Lastly, Section 5 concludes the paper.

2. An Overview of Mortality Estimation by Survival Analysis

Survival analysis has been developed in the field of bio-statistics to describe the timing of events. It has become a subject of increasing interest to health economists. For example, in Taiwan, Zimmer, Martin, and Lin (2005) analyzed the determinants of old-age mortality in Taiwan. The data used was from the Survey of Health and Living Status of the Elderly in Taiwan (SHLS). Initial interviews in 1989 with 4049 respondents aged 60 and over were followed up in 1993 and 1996. Data were linked to a death register that provides the exact date of death for those who died beginning at the day of initial interview until 1999, thus providing a comprehensive information to examine socio-demographic and health status differentials in subsequent mortality. The estimation methods were used the Gompertz regression model and both fixed and time-varying covariates. They found that functional and global assessments of health have stronger association with survival than reports of other health-related characteristics. Mainlanders have higher survival than others.

Lin and Lin (2006) employed the Cox proportional hazard model to investigate the factors associated with survival status of the elderly in Taiwan. Their specification of the model solved the impact of factors on the survival status of older Taiwanese during the period of 1989 to 1999. The data in their empirical study was also obtained from the SHLS. They found that ten factors, including age, gender, ethnicity, marital status, social participation, self-rated health, smoking, consumption of areca nuts, ADL function, and physical function had a good prediction index for survival status of the elderly in Taiwan.

Ho (2008) also introduced the Cox proportional hazard model to analyse the influence of living arrangements and health care utilizations on total mortality among the middle aged and elderly in Taiwan. Subject data was also obtained from the SHLS, but the study conducted in 1996 that encompassed observations on 2,462 individuals aged 50 and over. Survey data was linked to 1996-2003 national death registry data. Principal empirical results confirmed that, after controlling for potentially confounding variables, the relatively younger elderly had a higher survival rate during the period 1996 to 2003. Females also had a longer life span than males in the same period.

The above studies, however, did not consider the influence of unobserved heterogeneity on estimated mortality, they might produce incorrect results (Lancaster, 1990). Therefore, this paper extends the specified data from Ho (2008), and uses the Weibull model to fill this gap. The expected result confirm that, after considering unobserved heterogeneity, most estimated coefficients on the frailty model are larger in magnitude that the corresponding coefficients in the non-frailty model.

3. Methods

3.1 Data Source

The data used is from the Survey of Health and Living Status of the Middle Aged and Elderly in Taiwan (SHLS), a joint survey conducted by the Taiwan Provincial Institute of Family Planning and the Population Studies Centre, University of Michigan. The SHLS data has three panels. This paper only uses the second panel, which was conducted between 1996 and 2003. Initial interviews were held in 1996 with 2462 respondents aged 50 and over. Follow up interviews were conducted in 1999 and 2003. By the end of the 7-year period, 252 original respondents had died. This paper tracks the survival of the 2462 respondents over the 7-year period and analyzes the relative determinants of mortality.

3.2 Variables Specification

3.2.1 Dependent Variable

According to the SHLS data, the sample consists of two groups, including those currently alive and those who had died during the observation period. The former was designated as “right-censored” observation periods. The latter was known as “uncensored” observation periods. For “uncensored” observation periods, the study duration began with the dates when individuals were first interviewed and ended with the dates of death. Conversely, “right-censored” observation periods were alive throughout the study period. This variable can be categorized as dependent, with the uncensored variable coded 1 for deceased and 0 for still living.

3.2.2 Explanatory Variables

Multiple stepwise regression analysis was applied to determine variables that were significant predictors of mortality hazard. The explanatory variables included age, gender, living arrangement, and health care utilization factors. First, the age variable can be categorized to four groups: Age 1 (aged 50 to 54), Age 2 (aged 55 to 59), Age 3 (aged 60 to 64), and Age 4 (aged 65 and over). Age 1 was designated as a reference variable. Age categories might correspond to the retirement ages of 50, 55, 60, or 65 years, as designated in Taiwan’s Labor Standard Law.

Next, the gender variable was coded 1 for women and 0 for men. Based on the results of previous studies, that relatively older persons and males might expect to be more likely to die than relatively younger persons and females. In terms of living arrangement variables, it can be classified five possible living arrangements, namely living alone, living with spouse only, living with children only, living with children and spouse, and living with others. Living with others was defined as elderly subjects living with relatives, friends, or in social welfare institutions. The “living alone” was used as the reference. Finally, in terms of health care utilization variables, this paper considers Chinese and western medicine services simultaneously. It can be categorized results along three variables, namely days of staying in hospital during the past year (1 week, 2 weeks, 15-30 days, and 30 or more days), numbers of clinic visits (1-3, 4-6, 7-9, and 10 or more) and number of medicines bought (1-2, 3-4, and 5 or more) during the one month period immediately preceding the interview. Those who had not been hospitalized, visited a clinic or purchased medicine during this period were the references respectively. This study expects that survival rates correlate negatively with days spent in the hospital, number of doctor visits and number of medicines purchased.

3.3 Weibull Models without Unobserved Heterogeneity

Weibull distributions are widely used as models for survival analysis. The hazard function without unobserved heterogeneity is specified as

$$h(t \mid x_i) = \alpha t^{\alpha-1} \cdot \lambda = \alpha t^{\alpha-1} \cdot e^{(\beta_0+\beta_i x_i)}. \tag{1}$$

Empirically, the parameters λ and α in the Weibull distribution can be estimated by maximum likelihood. The parameter λ depends on the explanatory variables x_i, thus providing us with a more flexible hazard function. For example, the hazard function is increasing if $\alpha > 1$, decreasing if $\alpha < 1$, and constant if $\alpha = 1$. For observed duration data, $t_1, t_2, ..., t_n$ the log-likelihood function can be formulated and maximized to include censored and uncensored observations. Combining these survival models into a general parametric likelihood yields:

$$L(\beta) = \prod_{i=1}^{n} \left\{ \left[f(t_i | x_i, \beta) \right]^{c_i} * \left[S(t_i | x_i, \beta) \right]^{1-c_i} \right\}. \tag{2}$$

where $\beta = (\lambda, \alpha)$, and $c_i = 1$ represents uncensored observations, $c_i = 0$ represents right-censored observations (Cleves, et al, 2002). To obtain the maximum likelihood with respect to the parameters of interest, β, then maximise the log-likelihood function:[1]

$$\ln L(\beta) = \sum_{i=1}^{n} \left\{ c_i \ln\left[f(t_i | x_i, \beta) \right] + (1 - c_i) \ln\left[S(t_i | x_i, \beta) \right] \right\}. \tag{3}$$

The procedure to obtain the values of maximum likelihood estimation requires taking derivatives of $\ln L(\beta)$ with respect to β, the unknown parameters, setting these equations equal to zero, and solving for β.[2]

3.4 Weibull Models with Unobserved Heterogeneity

After considering unobserved heterogeneity on estimated mortality, the hazard function by frailty model can be defined as

$$h(t \mid x_i, u) = \alpha t^{\alpha-1} \lambda = \alpha t^{\alpha-1} \cdot e^{(\beta_0+\beta_i x_i + u)}. \tag{4}$$

where u can represent unobserved heterogeneity, the differences between observations are introduced via a multiplicative scaling factor. This is a random variable taking on positive values, with the mean normalized to one and finite variance σ^2. A crucial assumption in the model is that u is distributed independently of x_i and t. The same previous procedure to obtain the values of maximum likelihood estimation requires taking derivatives of $\ln L(\beta)$ with respect to β, the unknown parameters, setting these equations equal to zero, and solving for β.

---

[1] Since the log function is monotone, maximum of (2) and (3) occur at the same value of β; however, maximizing (3) is computationally simpler than maximizing (2).

[2] See Klein and Moeschberger (1997), for a description of the numerical methods for implementing multivariate Newton-Raphson methods.

4. Results

Frailty is a random component designed to account for variability due to unobserved individual-level factors that is otherwise unaccounted for by the other predictors in the Weibull model. In particular, suppose the SHLS data belonging to a random sample, the shared frailty model by gender can be used for examining the effects of unobserved heterogeneity on estimated morality among the middle aged and elderly in Taiwan.

The empirical results are shown in Table 2. First, before considering unobserved heterogeneity, the estimated coefficients of those with Age3 (aged 60 to 64) and Age4 (aged 65 to 70) variables are positive and statistically significant. This means that older people have higher hazard rates of mortality *ceteris paribus*. Further, in terms of health care utilization variables, the estimated coefficients of those with Hosp3 (Staying in hospital 8-14 days), Hosp4 (Staying in hospital 15-30 days), Hosp5 (Staying in hospital 31 and over days), and Clinic5 (Clinic visiting 10 and over times) variables have a positive and statistically significant effect on mortality. This implies that those people have higher hazard rates of mortality *ceteris paribus*.

In contrast, in terms of living arrangements, the estimated coefficients of those with Spouse (Living with spouse), Children (Living with children), Both (Living with spouse and children), Clinic4 (Clinic visiting 7-9 times), Bmedic2 (Buying medicines 1-2 times), and Bmedic3 (Buying medicines 3-4 times) variables are negative effect on mortality, but only the variables of Both and Bmedic2 are statistically significant. This means that people living with spouse and children, and those who buying medicines 1-2 times have lower hazard rates of mortality *ceteris paribus*. Perhaps, they have better living arrangements and easiler enjoy their later livies. The estimate for the shape parameter is 1.084 suggesting an increasing hazard rate of mortality over time.

Second, after considering unobserved heterogeneity, the frailty model is assumed to follow a gamma distribution with mean 1 and variance equal to theta(θ). The estimate of theta is 0.086. A variance of zero (theta = 0) would indicate that the frailty component does not contribute to th e model. A likelihood ratio test for the hypothesis theta = 0 is shown directly below the parameter estimates and indicates a chi-square value of 14.48 with 1 degree of freedom yielding a highly significant p-value of 0.000. Notice how all the parameter estimates are altered with the inclusion of the frailty. The estimate for the Weibull distribution shape parameter is now 1.087, different from the estimate 1.084 obtained from the model without frailty. The inclusion of frailty not only has an impact on the parameter estimates but also complicates their interpretation.

Interestingly, the estimated coefficients of Age3 (aged 60 to 64) and Age4 (aged 65 to 70) are a little bit smaller in magnitude that the corresponding coefficients in the reference model. This suggests that the elderly might less consider the effect of unobserved heterogenity on estimated mortality. In contrast, in terms of health care utilization variables, the estimated coefficients on the regressors Hosp3 (Staying in hospital 8-14 days), Hosp4 (Staying in hospital 15-30 days), Hosp5 (Staying in hospital 31 and over days), and Clinic5 (Clinic visiting 10 and over times) are a little bit larger in magnitude that the corresponding coefficients in the reference model. This implies that the terms of health care utilization have larger unobserved heterogenity on estimated mortality. The hospital managers need to consider how to control and decrease the effect of unobserved heterogenity, and provide better services for the patients.

Furthermore, in terms of living arrangements, the estimated coefficients of those with Spouse (Living with spouse), Children (Living with children), Both (Living with spouse and children) variables are a little bit larger in magnitude that the corresponding coefficients in the reference model, but it is insignificant. This suggests that the terms of living arrangements might need more data and time to analyse the effct of unobserved heterogenity on estimated mortality. Finally, the Weibull distribution shape parameter α is also slightly larger in the frailty model than in the reference model.

5. Conclusion

Using the SHLS, this paper provides evidences on the effects of unobserved heterogeneity on estimated mortality among the middle aged and elderly in Taiwan between 1996 and 2003. The mortality hazards are estimated by Weibull models. Main results confirm that, after considering unobserved heterogeneity, most estimated coefficients on the mortality hazard are larger in magnitude that the corresponding coefficients in the reference model. Especially, the terms of health care utilization have larger unobserved heterogenity on estimated mortality. Further, the variables of ages are a little bit smaller in magnitude that the corresponding coefficients in the reference model. Therefore, if the government policies can concern more unobserved heterogenity of health care utilization that might useful to decrease the mortality for the elderly.

Due to data limitations, this paper may not be generalized for mortality associated with diseases and their symptoms. In particular, different illnesses and morbidity rates also have some effects of unobserved heterogenity. In the future, the authors plan to examine more carefully the influence of unobserved heterogenity by disease and symptoms indicators on estimated mortality.

References

Cleves, M. A., Gould, W. W., and Gutierrez, R. G. (2002). *An Introduction to Survival Analysis Using Stata,* Stata Corporation, USA.

Ho, S. (2008). “Survival Analysis of Living Arrangements and Health Care Utilization in Terms of Total Mortality Among the Middle Aged and Elderly in Taiwan” *The Journal of Nursing Research*, 16(2), 1-9.

Hosmer, D. W., and Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modelling of Time to Event Data*. John Wiley & Sons, Inc.

Klein, J. P., and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag. New York, Inc.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data.* Cambridge University Press.

Lin, H. C., & Lin, H. S. (2006). “Factors associated with survival status of older Taiwanese adults” *Taiwan Public Health Association,* 25(5), 351-362 (in Chinese).

Zimmer, Z., Martin, L. G., & Lin, H. S. (2005). “Determinants of old-age mortality in Taiwan” *Social Science and Medicine, 60*(3), 457-470.

Table 1 Descriptive Statistics of Variables

Variables	Description	Mean	Standand Error
DURATION	1-92 months.	85.752	(15.980)
CENSOR	1 = Uncensored, 0 = Otherwise.	.102	(.303)
AGE1	1= Aged 50 to 54, 0 = Otherwise.	.286	(.452)
AGE2	1 = Aged 55 to 59, 0 = Otherwise.	.325	(.469)
AGE3	1 = Aged 60 to 64, 0 = Otherwise.	.269	(.444)
AGE4	1 = Aged 65 to 70, 0 = Otherwise.	.119	(.324)
GENDER	1 = Female, 0 = Male.	.485	(.499)
SINGLE	1 = Living alone, 0 = Otherwise.	.055	(.228)
SPOUSE	1 = Living with spouse, 0 = Otherwise.	.122	(.328)
CHILDREN	1 = Living with children, 0 = Otherwise.	.140	(.347)
BOTH	1 = Living with spouse and children, 0 = Otherwise.	.680	(.467)
HOSP1	1 = Never stay in hospital, 0 = Otherwise.	.866	(.341)
HOSP2	1 = Stay in hospital 1-7 days, 0 = Otherwise.	.060	(.237)
HOSP3	1 = Stay in hospital 8-14 days, 0 = Otherwise.	.028	(.166)
HOSP4	1 = Stay in hospital 15-30 days, 0 = Otherwise.	.027	(.163)
HOSP5	1 = Stay in hospital 31 and over days, 0 = Otherwise.	.019	(.135)
CLINIC1	1 = Never clinic visiting, 0 = Otherwise.	.483	(.499)
CLINIC2	1 = Clinic visiting 1-3 times, 0 = Otherwise.	.377	(.485)
CLINIC3	1 = Clinic visiting 4-6 times, 0 = Otherwise.	.066	(.248)
CLINIC4	1 = Clinic visiting 7-9 times, 0 = Otherwise.	.006	(.080)
CLINIC5	1 = Clinic visiting 10 and over times, 0 = Otherwise.	.068	(.252)
BMEDIC1	1 = Never buying medicine, 0 = Otherwise.	.710	(.454)
BMEDIC2	1 = Buying medicines 1-2 times, 0 = Otherwise.	.183	(.387)
BMEDIC3	1 = Buying medicines 3-4 times, 0 = Otherwise.	.044	(.206)
BMEDIC4	1 = Buying medicines 5 and over times, 0 = Otherwise.	.063	(.244)

Note:

The sample has 2462 observations, including 252 death (event observations) and 2210 currently alive (right-censored observations).

Table 2 Mortality Estimation by Weibull Model

	Without Unobserved Heterogeneity		With Gamma- Heterogeneity	
Variables	Coefficient	Standard Error	Coefficient	Standard Error
AGE2	.206	(.200)	.212	(.200)
AGE3	.866***	(.189)	.865***	(.189)
AGE4	1.052***	(.210)	.995***	(.211)
SPOUSE	-.300	(.264)	-.167	(.266)
CHILDREN	-.305	(.255)	-.037	(.264)
BOTH	-.400*	(.217)	-.324	(.218)
HOSP2	.381	(.244)	.390	(.244)
HOSP3	.796***	(.280)	.800***	(.281)
HOSP4	1.319***	(.243)	1.323***	(.242)
HOSP5	1.900***	(.244)	1.949***	(.245)
CLINIC2	.194	(.145)	.228	(.146)
CLINIC3	.178	(.244)	.228	(.245)
CLINIC4	1.376	(.341)	1.762	(.719)
CLINIC5	.655***	(.211)	.762***	(.213)
BMEDIC2	-.541***	(.203)	-.570***	(.203)
BMEDIC3	-.159	(.291)	-.163	(.291)
BMEDIC4	.256	(.228)	.262	(.229)
Constant	-7.539***	(.399)	-7.674***	(.450)
$/\ln\_\alpha$	.080	(.061)	.084	(.061)
$/\ln\_the$			-2.450**	(1.089)
α	1.084	(.066)	1.087	(.066)
$1/\alpha$	.923	(.056)	.920	(.056)
theta			.086	(.094)
No. of sample	2462		2462	
No. of death	252		252	
Log likelihood	-984.808		-977.566	
LR chi2 (17)	147.65***		150.97***	

Notes:

1. Effects are significant at * $p \le .10$, ** $p \le .05$, *** $p \le .01$.
2. Group variable is gender, including 1268 men and 1194 women.
3. Goodness of fit: the result of Log-likelihood ratio test can reject the hypothesis that all coefficients except the intercept are 0 at the 0.01 level. In particular, Log-likelihood ratio test of theta = 0: chibar2 (01) = 14.48, Prob >= chibar2 = 0.000.