ANALYZING THE EFFECT OF HOUSEHOLD INCOME, PARENT DEATH AND SPOUSE DEATH ON SURVIVAL OF RETIRED PEOPLE: A LEFT TRUNCATED MARGINAL COX MODEL WITH SIMEX CORRECTION FOR MEASUREMENT ERROR

by

YE JIN

(Under the Direction of Xiao Song)

ABSTRACT

We aimed to investigate the associations and to estimate the effects of the parent's death, spouse's death, and household income on survival time of retired people. We used left truncated right-censored marginal Cox regression model on HRS dataset in our study. We also implemented the simulation and extrapolation (SIMEX) method in SAS program, and used this method to correct the effect of measurement error of household income. We then applied the bootstrap approach to obtain the standard error and the 95% confidence interval for the SIMEX estimates. We found that the mortality risk is significantly associated with mother's death, father's death, spouse's death, and household income. Parent's death and spouse's death both have negative impacts on mortality risk. The mortality risk would be less, if an individual has higher household income. The measurement error does not change the effects of these variables significantly.

INDEX WORDS: LEFT TRUNCATION, MARGINAL COX MODEL, SIMEX, HOUSEHOLD INCOME, PARENT'S DEATH, SPOUSE'S DEATH

ANALYZING THE EFFECT OF HOUSEHOLD INCOME, PARENT DEATH AND SPOUSE DEATH ON SURVIVAL OF RETIRED PEOPLE: A LEFT TRUNCATED MARGINAL COX MODEL WITH SIMEX CORRECTION FOR MEASUREMENT ERROR

by

YE JIN

B.S., University of Georgia, 2017

A Thesis Submitted to the Graduate Faculty
of The University of Georgia in Partial Fulfillment
of the
Requirements for the Degree
MASTER OF SCIENCE

ATHENS, GEORGIA

2019

ANALYZING THE EFFECT OF HOUSEHOLD INCOME, PARENT DEATH AND SPOUSE DEATH ON SURVIVAL OF RETIRED PEOPLE: A LEFT TRUNCATED MARGINAL COX MODEL WITH SIMEX CORRECTION FOR MEASUREMENT ERROR

by

YE JIN

Major Professor:	Xiao Song
Committee:	Kevin K. Dobbin
	Stephen L. Rathbun

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2019

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUTCTION

1.1 BACKGROUND

Parents play an important role in an individual's life. Loss of a parent could dramatically impact an individual's health condition (Pennebaker, 1985), and could also increase the risk of an individual's mortality (Rostila & Saarela, 2011). Rostila & Saarela (2011) conducted a cohort study in Swedish and identified an effect of early-life parent death on all-cause mortality of offspring. Saarela & Rostila (2018) found a significant association between the death of a parent in adulthood and all-cause mortality in a cohort study performed in Finland. The estimated association in Finland is larger than that in Sweden. Li et al. (2014) also identified a positive effect of parental death in childhood or adolescence on the risk of all-cause mortality in early adulthood based on the data of a cohort study in Denmark. Another study focused on the effect of loss in early life on later life mortality. It used questionnaires to collect necessary data, and found a significant association between loss in early life and later life mortality in the population of Utah (Smith et al., 2014). Therefore, both early life and adulthood parental loss could affect the risk of all-cause mortality.

Spouses have a strong connection with each other. A study in Israel shows evidence of the association between a spouse's death and excess mortality. The risk of mortality increases especially in a short duration after the death of a spouse (Manor & Eisenbach, 2003). A Finnish study shows that the risk of mortality increases after the death of a spouse. This study also shows that excess mortality after a spouse' death is partially caused by the loss of social support or stress (Martikainen & Valkonen, 1996). Helsing et al. (1981) used the information collected at Maryland to study factors associated with mortality after widowhood. Besides an interaction between the death of a spouse and gender on mortality rate, the author also found living alone could increase the mortality rate. Comparing with living in or moving between residents, moving into a nursing home or other chronic care facility could also increase the risk of mortality. Hence, there is an association between the death of a spouse and excess mortality.

The financial condition can influence the lifestyle for individuals, and can also influence the mortality rate. Kaplan et al. (1996) conducted a study in the United States and found that income inequality is significantly associated with health outcomes and mortality rate. Backlund et al. (1996) studied the relationship between income and mortality using national wised data. The author identified the association between income level and mortality rate. The author also found that compared with elders, the income-mortality gradient for working-age adults is larger. Demakakos et al. (2015) investigated the relationship between wealth and mortality for people over age 50 and found that wealth is significantly associated with mortality. Thus, financial conditions would be important risk factors of mortality.

Retirement is a changing point for most people. The physical and mental condition, family relationship, economic welfare, and social support change dramatically before and after retirement. Limited studies focus on the relationships between mortality and the death of parents, death of a spouse, and financial condition for retirement and older age people. Hence, studies are needed to investigate this effect.

1.2 HRS DATA OVERVIEW

The Health and Retirement Study (HRS) is a nationally representative cohort study of individuals in the USA. The HRS represents Americans over age 50. The study follows individuals and their spouse from the time they enter the survey until death. Each individual is identified by a household ID and a personal ID, and the unique identifier for each participant is the combination of the household ID and personal ID. The HRS study began to recruit participants in 1992, and the sample is built over time. A survey was taken every 2 years. Besides basic information, the HRS study mainly focuses on four topics – financial condition; health condition and healthcare usage; work and retirement; and family connection. The response rate for each cohort is greater than 80% (Sonnega et al., 2014). We collected the death time from RAND HRS fat files from 1992 to 2014, which contain all the raw variables. We use a subset of RAND HRS longitudinal file 2014 (v2), which is a cleaned version of the HRS dataset ("Health and Retirement Study Data File", no date). The subset contains 27965 observations and 18146 households, with information of race, gender, birthday, death time, mother's death time, father's death time, spouse's death time and total household income (response & spouse) for maximum 12 follow-ups. We calculated survival event time using death time and birthday, and recorded in years. We used the response indicator and birthday to calculate the age of entry study, and the household income was repeatedly measured during the follow-up. We take the expected value of household income, which can give us an idea of the measurement for the household income. We also calculated the age at the time of their parents and spouse death.

1.3 PURPOSE OF STUDY

We aimed to investigate the associations and to estimate the effects of parents' death, spouses' death, and household income on survival time.

A popular approach to analysis (right-censored) survival data is the Cox proportional hazards model (Cox, 1972). Several complications are encountered in the analysis of the HRS data. First, there might be a correlation between the survival times of couples. Second, the survival time is left-truncated. Third, there might be a measurement error on household income.

1.4 LITERATURE REVIEW

Survival data can be divided into clusters by interpersonal relationships (couples, siblings, and families), communities, and geographical regions. People in the same cluster tend to be correlated because they usually share certain characteristics. The cluster effect is often ignored in most analyses. The ignorance of clusters may lead to unpredicted bias and inefficiency. Two different approaches focused on this issue are the frailty model and the marginal model. The frailty model adopts random effects to characterize the correlation among clusters (McGlichrist, 1993). One restriction of the frailty model is that it requires a parametric distribution for the random effect. For example, in the gamma frailty model, the cluster effect, which is frailty, is assumed to follow a gamma distribution. In contrast, the marginal model (Lin, 1994) does not impose a specific dependence structure among the clusters. Robust standard error estimation is used to account for the intra-cluster correlation.

It is common to consider survival data as left-truncated data, when the event time is the lifetime of an individual while he was not observed from the time of birth. There will be bias introduced if left truncation is not considered, because prior events are not observed. The risk set for left-truncated data is different compares to the risk set without truncation (Klein & Moeschberger, 2005).

Measurement error is a common problem for data analysis, especially in observational data. It will introduce bias in estimation and lead to power loss. There are a large number of methods with different assumptions aiming to correct for measurement error (Carroll et al., 2006). These methods could be categorized into several kinds of approaches: regression calibration method, likelihood based approaches, estimating equation based on approaches and simulation extrapolation method. Suppose X is the vector of true predictor variables. The regression calibration method is an approximate method. It reduces bias but does not remove bias completely. The likelihood approaches usually require parametric distributions on X and the errors. The estimating equation based on approaches such as the conditional score and corrected score approaches do not require distribution assumption on X and/or the error, and can be more robust. Simulation-extrapolation (SIMEX) method is a general approach to deal with measurement error and it is easy to implement. It does not require a specific distribution of X.

CHAPTER 2

METHODOLOGY

2.1 COX MODEL

We use Cox model (Cox, 1972) to study the relationship between the time to event and covariates. For subject *i*, let T_i denote the survival time, X_i be a vector of *k* covariates, the hazard rate at time *t* can be written as:

$$h(t|X_i) = h_0(t)\exp(\beta^T X_i),$$

where $h_0(t)$ is an unspecified baseline hazard function. Suppose the event time is ordered as $t_1 < t_2 < \ldots < t_D$, and denote $R(t_r)$ the risk set at event time t_r, and i_r be the subject died at t_r. Based on the hazard function above, the partial likelihood is

$$L(\beta) = \prod_{r=1}^{D} \frac{\exp(\beta^T X_{i_r})}{\sum_{j \in R(t_r)} exp(\beta^T X_j)}.$$

The denominator of likelihood depends on the information of all individuals that have not experience the event yet, and the numerator only depends on the information of individuals with the event occurred (Klein & Moeschberger, 2005).

2.2 LEFT TRUNCATION

Since the HRS study focuses on retired people, individuals who died earlier are not included in the study, which caused left truncation of the survival time. Individuals are only observed after they entered the study. It is well known that age is highly correlated with mortality rate. Thiebaut et al. (2004) showed for left-truncated data, even the effect of age is controlled, bias could still be introduced into the study if we use time-on-study as the time scale instead of age. Hence, it is important to adjust for left-truncation in analyzing the HRS data. A conditional distribution should be used in likelihood construction. Specifically, the partial likelihood is modified to accommodate the delayed entry in the risk set, which is redefined as $R(t) = \{j: V_j \le t \le T_j\}$, where V_j is the age that the person entered the study (Klein & Moeschberger, 2005).

2.3 THE MARGINAL APPROACH OF COX MODEL

When we study the HRS dataset, we need to consider that there are individuals coming from the same household. There might be a correlation between individuals if they belong to the same household. We consider each household as a cluster, and the marginal approach of Cox regression is used to deal with the correlation within the same cluster. For the *ith* individual in the *jth* group at time *t*, let X_{ij} denote the covariates, the marginal Cox regression model assumes that the hazard

$$h\left(t \middle| X_{ij}\right) = h_0(t)\exp\left(\beta^T X_{ij}\right).$$

Inference based on the marginal model takes into account of the correlation within the clusters. Comparing the standard Cox model with the marginal approach of the Cox model, the parameter estimations for both models are the same. However, the naive variance estimates do not reflect the variance-covariance structure of the clustered data and do not account for the intra-group correlation, which will yield a loss of power on the effect estimation. Hence, the robust variance estimate proposed by Lee et al. (1992) is used in the marginal approach of the Cox model.

2.4 SIMULATION EXTRAPOLATION

The household income (in dollars) was obtained from the questionnaire, which may subject to measurement error. Ignoring the measurement error might lead to biased estimation of the regression coefficients and incorrect inference (Carroll et al., 2006) (Oh, Shepherd, Lumley & Shaw, 2017). To reduce the bias due to measurement error, we adopt the simulation extrapolation (SIMEX) method. SIMEX is a simulation-based method. It is implemented in two steps, the simulation step and the extrapolation step, which are described as follows. In the resampling stage, it adding additional measurement error to the data, and obtains the trend of estimated coefficient change versus measurement error. Then, extrapolate the trend-estimated coefficient to obtain the estimate in the case of no measurement error. Suppose X_i is the predictor without measurement error for the *ith* subject, and $W_{ij} = X_i + U_{ij}$ $(j = 1,\ldots, m_i)$ be the measured value of X_i, U_{ij} is a normal random variable with mean zero and variance σ_U^2.

Step 1: Simulation Step

The simulation step creates datasets with additional measurement error. Consider a sequence of λ, $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_q$. For each $\lambda > 0$, we generated B simulated datasets. Specifically, for $b = 1, 2, \ldots, B$, let

$$W_{bi}(\lambda) = \overline{W_i} + \sqrt{\lambda}\sigma_i Z_i, \text{ with } \sigma_i = \sigma_U^2 / \sqrt{m_i},$$

where Z_i is generated from the standard normal distribution (Carroll, 2006). Then, we fit the model using the simulated data, and get the estimated coefficients $\hat{\beta}_b(\lambda)$.

Step 2: Extrapolation Step

Let $\bar{\beta}(0)$ be the naïve estimate based on the original dataset. For each $\lambda > 0$, we take the average of estimated coefficients $\hat{\beta}_b(\lambda)$ ($b = 1, 2, \ldots, B$) and obtain

$$\bar{\beta}(\lambda) = \frac{\sum_{b=1}^{B} \hat{\beta}_b(\lambda)}{B}.$$

The SIMEX estimate of β is obtained by extrapolation of the estimated $\bar{\beta}(\lambda)$ to $\lambda = -1$ base on a regression model for each coefficient. A quadratic model is usually used for the extrapolation.

2.5 BOOTSTRAP

We use bootstrap to estimate the standard error of the SIMEX estimator. We generated 50 independent bootstrap samples. We use the %boot macro (http://support.sas.com/kb/24/982.html) in SAS, which is created by SAS to implement the bootstrap procedure, and obtain the SIMEX estimates for each bootstrap dataset. Based on these results, we calculate the standard error of the SIMEX estimates and the 95% confidence interval of the regression coefficients.

CHAPTER 3
APPLICATION

3.1 DATA DESCRIPTION

For the HRS dataset, we would like to evaluate the effect of the parent's death, spouse's death, log transformed household income, birthplace, gender, and race on survival time. The summary statistics of the covariates are shown in table 1.

Table 1: Baseline Statistics

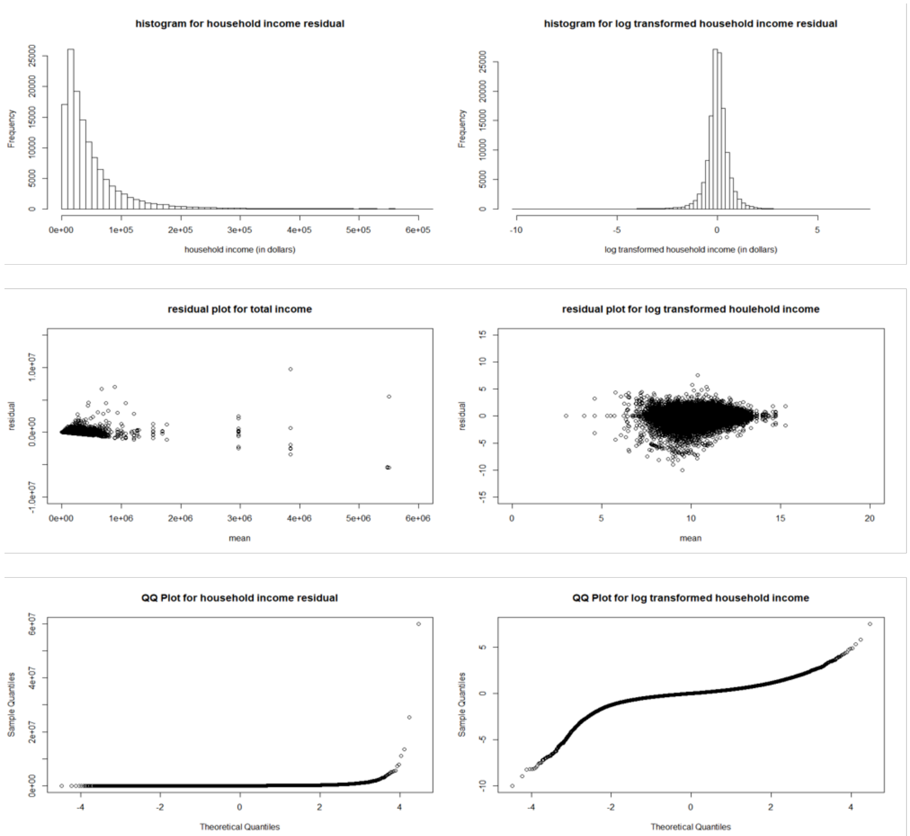
Variable	Total
Place of Birth*	
0: outside of US, N (Percentage)	3907 (13.97%)
1: in US, N (Percentage)	24031 (85.93%)
Gender	
1: Male, N (Percentage)	11831 (42.31%)
2: Female, N (Percentage)	16134 (57.69%)
Race*	
1: Other, N (Percentage)	2195 (7.85%)
2: Black, N (Percentage)	5152 (18.42%)
3: White, N (Percentage)	20547 (73.47%)
Mother's Death	
0: Alive, N (Percentage)	15283 (54.65%)
1: Death, N (Percentage)	12682 (45.35%)
Father's Death	
0: Alive, N (Percentage)	15004 (53.65%)
1: Death, N (Percentage)	12961 (46.35%)
Spouse's Death	
0: Alive, N (Percentage)	23859 (85.32%)
1: Death, N (Percentage)	4106 (14.68%)
Age, Mean (SD)	57.36 (8.91)
Log transformed household income, Mean (SD)	10.36 (0.97)

** Variables that contain missing value*

The total number of participants is 27965. Among these participants, there are 5275 observed deaths during the study period. 85.93% of the participants were born in the U.S, 57.69% are females. There are 73.47% white people, 18.42% are black, and 7.85% are from other races. There are 12682 (45.35%) participants with observed mother's death, 12961 (46.35%) with observed father's death and 4106 (14.68%) with observed spouse's death. The mean of log transformed household income is 10.36, and the standard deviation is 0.97.

3.2 MODEL APPLICATION

We use the left-truncated marginal Cox model to analysis the HRS dataset. We calculate the age of individuals (in years) at the losses of their parents or spouse, entry of the study, and death. The household income (in dollars) were observed once or multiple times during the study. There are 18146 households, and 27965 participants. Hence, many households contain more than one participant, and there is a large number of small groups. There are 333 households with more than two subjects, which may cause by remarriage. When we consider the death of a spouse, we only record the first spouse's death.



a, histogram for household income residual; d, histogram for log transformed household income residual;
b, residual vs mean plot for total income; e, residual vs mean plot for log transformed total income;
c, QQ plot for household income residual; f, QQ plot for log transformed household income residual.

Figure 1: Plots for Household Income Residual Before and After Log Transformation

Since the household income (in dollars) was obtained from the questionnaire, we should consider measurement error. To apply the SIMEX method, we first check if the household income has a normally distributed error. Let W_{ik} be the kth observation *(k=1,…,*m_i*)* of household income for the ith household. Figure 1(a) shows the histogram of the residuals ($W_{ik} - \overline{W_i}$), and it is skewed to the right. Figure 1(b) is the residual ($W_{ik} - \overline{W_i}$) vs. mean ($\overline{W_i}$) plot which shows an apparent heteroscedasticity pattern. The Q-Q plot of the residuals in figure 1(c) also shows an apparent skewed pattern. To achieve approximately within subject normality and constant variance, we apply log transformation to household income. Figure 1(d) is the histogram of the residuals after the log transformation, and it shows an approximately bell shape. The residual vs. mean plot (Figure 1(e)) and the residual Q-Q plot (Figure 1(f)) show that after log transformation the residual is approximately normally distributed.

Parents' death and spouse's death are potential confounding factors of the individuals' survival time. We considered the mother's death and father's death separately. Since the parents' death status and the death status of a spouse are changing over time, we include three time-dependent covariates for the mother's death, father's death, and the death of a spouse. Let $M_1(t) = I(mother\ died\ by\ time\ t)$, $M_2(t) = I(father\ dies\ by\ time\ t)$, and $M_3(t) = I(spouse\ died\ by\ time\ t)$, where I is the indicator function. The indicator will change from 0 to 1 when the corresponding death happened. Denote the true log household income by X_i and other covariates by $Z_1, \dots, Z_r$. For individual *j* in household *i*, the marginal Cox model can be written as

$$h(t|X_i, M_{ij}(t), Z_{ij}) = h_0(t)\exp\{\beta_1 X_i + \sum_{p=1}^{3} \beta_{p+1} M_{pij}(t) + \sum_{p=1}^{r} \beta_{p+4} Z_{pij}\}.$$

950 observations were deleted from the model due to the missing value.

We analyzed data using both the naïve approach and the SIMEX approach. For the SIMEX approach, we take $\lambda = 0, 0.2, 0.4, \dots, 2$, and $B = 100$. The estimated measurement error variance for log transformed household income is 0.401, which is obtained by

$$\widehat{\sigma_U^2} = \frac{\sum_{i=1}^{h}\sum_{k=1}^{mi}(W_{ik} - \overline{W_i})^2}{\sum_{i=1}^{h}(m_i - 1)}.$$

3.3 RESULTS

The SIMEX estimates are obtained from simulation and extrapolation. Figure 2 shows the extrapolation of $\bar{\beta}(\lambda)$ to $\lambda = -1$ to obtain the SIMEX estimates. In each plot, $\bar{\beta}(\lambda)$s are close to the fitted quadratic curve.

Figure 2: Plots of $\bar{\beta}(\lambda)$ vs. λ with the overlaid fitted curve in Extrapolation Step.

Table 2 shows the estimated coefficients for the standard Cox model, marginal approach, and SIMEX approach with white as the reference group for race, male as the reference for gender, and born outside of the U.S. as the reference of birthplace. We fit the model using both datasets with and without households contains more than two individuals, and the results are similar. Hence, we will not delete households with more than two individuals from the dataset.

The naïve estimates based on the standard Cox model and marginal Cox model are the same. There are slight differences between the standard errors for the standard Cox model and the standard errors for the marginal Cox model. Comparing with the standard Cox model, race(other), mother's death, father's death, and log transformed household income have a slightly larger standard error in the marginal Cox model. Birthplace has a slightly smaller standard error in the marginal Cox model. For spouse's death and race (black), the standard error is about the same for both the standard Cox model and the marginal Cox model.

The coefficient estimates after correcting the measurement error on log transformed household income using SIMEX are slightly different. The estimated coefficients for log transformed household income, race, and gender are slightly smaller after correcting the measurement error. The estimated coefficient for spouse's death is about the same before and after SIMEX. After correcting the measurement error, the estimated coefficients for father's death and mother's death increase a little bit. The standard errors of the SIMEX estimates were obtained using 50 bootstrap datasets. Comparing to the marginal Cox model, the standard errors for birthplace and father's death decrease after SIMEX. The standard errors for other covariates increase after correcting the measurement error.

All covariates in the model have significant effects on survival time. Based on the result from the SIMEX approach, after adjusting for other covariates the hazard ratio for birthplace of born outside of the U.S. vs. in the U.S. is 1.3479. Comparing with people born outside the U.S., individuals born in U.S. is 1.3479 times more hazard. The estimated hazard of females is 0.6077 time of males. The estimated hazard ratio of black to white is 1.2278. There is no significant difference between white and other. Death of either parents or spouse will lead to the hazard increase. Mother's death, father's death, and spouse death will increase the hazard 1.4381, 1.1909, and 1.4428 times respectively. If the household income doubled, the hazard will decrease about a half.

Table 2: Result of Standard Cox model, Marginal Cox model, and SIMEX

with household contains more than two individuals

	Standard Cox Model						Marginal Cox Model						SIMEX					
Covariates	Cox model Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value	Marginal Cox model Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value	SIMEX Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value
birth place	0.2906	0.0525	1.3372	1.2040	1.4851	<.0001	0.2906	0.0509	1.3372	1.2078	1.4805	<.0001	0.2985	0.0474	1.3479	1.2259	1.4820	<.0001
gender (Female)	-0.4894	0.0292	0.6130	0.5783	0.6498	<.0001	-0.4894	0.0290	0.6130	0.5785	0.6496	<.0001	-0.4981	0.0346	0.6077	0.5671	0.6511	<.0001
race (Other)	0.0916	0.0800	1.0959	0.9339	1.2860	0.2524	0.0916	0.0840	1.0959	0.9264	1.2963	0.2757	0.0814	0.0916	1.0848	0.9032	1.3028	0.3742
race (Black)	0.2187	0.0392	1.2445	1.1506	1.3460	<.0001	0.2187	0.0410	1.2445	1.1465	1.3508	<.0001	0.2052	0.0529	1.2278	1.1045	1.3648	0.0001
mother's death	0.3605	0.0417	1.4340	1.3193	1.5586	<.0001	0.3605	0.0434	1.4340	1.3148	1.5640	<.0001	0.3633	0.0457	1.4381	1.3124	1.5759	<.0001
father's death	0.1727	0.0434	1.1885	1.0896	1.2964	<.0001	0.1727	0.0450	1.1885	1.0862	1.3005	0.0001	0.1747	0.0435	1.1909	1.0916	1.2992	<.0001
spouse's death	0.3653	0.0324	1.4410	1.3505	1.5375	<.0001	0.3653	0.0321	1.4410	1.3515	1.5364	<.0001	0.3666	0.0367	1.4428	1.3408	1.5527	<.0001
log transformed household income	-0.2167	0.0204	0.8052	0.7730	0.8388	<.0001	-0.2167	0.0219	0.8052	0.7708	0.8412	<.0001	-0.2397	0.0230	0.7869	0.7516	0.8239	<.0001

without household contains more than two individuals

	Standard Cox Model						Marginal Cox Model						SIMEX					
Covariates	Cox model Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value	Marginal Cox model Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value	SIMEX Parameter estimation	Standard Error	Hazard Ratio	Lower Bound	Upper Bound	p-value
birth place	0.2700	0.0529	1.3100	1.1784	1.4561	<.0001	0.2700	0.0514	1.3100	1.1820	1.4518	<.0001	0.2787	0.0590	1.3214	1.1742	1.4869	<.0001
gender (Female)	-0.4875	0.0296	0.6141	0.5788	0.6516	<.0001	-0.4875	0.0295	0.6141	0.5790	0.6514	<.0001	-0.4966	0.0333	0.6086	0.5694	0.6505	<.0001
race (Other)	0.0624	0.0825	1.0644	0.9025	1.2554	0.4492	0.0624	0.0862	1.0644	0.8959	1.2647	0.4687	0.0521	0.0851	1.0535	0.8887	1.2488	0.5404
race (Black)	0.2328	0.0396	1.2621	1.1659	1.3662	<.0001	0.2328	0.0416	1.2621	1.1614	1.3715	<.0001	0.2196	0.0408	1.2455	1.1480	1.3513	<.0001
mother's death	0.3680	0.0428	1.4448	1.3264	1.5739	<.0001	0.3680	0.0446	1.4448	1.3215	1.5797	<.0001	0.3705	0.0497	1.4485	1.3116	1.5998	<.0001
father's death	0.1720	0.0446	1.1877	1.0865	1.2984	0.0001	0.1720	0.0462	1.1877	1.0828	1.3028	0.0002	0.1742	0.0534	1.1903	1.0698	1.3243	0.0001
spouse's death	0.3727	0.0327	1.4516	1.3596	1.5498	<.0001	0.3727	0.0324	1.4516	1.3605	1.5488	<.0001	0.3738	0.0343	1.4533	1.3569	1.5565	<.0001
log transformed household income	-0.2137	0.0208	0.8076	0.7747	0.8419	<.0001	-0.2137	0.0223	0.8076	0.7723	0.8444	<.0001	-0.2380	0.0201	0.7882	0.7572	0.8205	<.0001

CHAPTER 4
DISCUSSION

In our study, we use the left truncated marginal Cox model to study the HRS dataset. According to the results, we found that for older adults, the risk of death is significantly associated with the death of mother, the death of father, the death of a spouse, and household income. Parent's death and spouse's death have a negative impact on mortality risk. The risk of death would be less, if an individual has higher household income. The measurement error does not change the effects of these variables significantly.

A Finnish study points out that the increased risk of mortality after the spouse's death is partially caused by stress (Martikainen & Valkonen, 1996). After bereavement people need to face a lot of stresses, both financial and emotional, and they need time to recover. Some people can also lose their social support after the death of a spouse. There are many reasons that can explain the association between the death of a spouse and excess mortality.

Secondly, parents' death also has effects on the risk of death. Parents always play an important role in a family. They are always the mental support of their children. The death of a parent or parents is a huge disaster in people's life, and will bring a lot of stresses. Financial stress, losing social support and emotional stress can all become problems at once. Parent's death has a negative impact on survival.

Thirdly, we studied about household income. Our result indicates that individual's with higher household income has a lower hazard of mortality. For people with higher income, they might have a better living condition and medical care.

We use the SIMEX method to correct measurement error on household income. Comparing the result before and after correcting measurement error, there is only a slight difference in both the estimation and the standard error for each covariate. Since all the covariates included in our data have really small p-values, there is no significant change after correcting the measurement error.

BIBLIOGRAPHY

Backlund, E., Sorlie, P., & Johnson, N. (1996). The shape of the relationship between income and mortality in the United States. *Annals Of Epidemiology, 6(1),* 12-20. doi: 10.1016/1047-2797(95)00090-9

Carroll, R. (2006). *Measurement error in nonlinear models.* Boca Raton: Chapman & Hall/CRC.

Cox, D. (1972). Regression Models and Life-Tables. *Journal Of The Royal Statistical Society: Series B (Methodological), 34(2),* 187-202. doi: 10.1111/j.2517-6161.1972.tb00899.x

Demakakos, P., Biddulph, J., Bobak, M., & Marmot, M. (2015). Wealth and mortality at older ages: a prospective cohort study. *Journal Of Epidemiology And Community Health, 70(4),* 346-353. doi: 10.1136/jech-2015-206173

Health and Retirement Study Data File. Retrieved from https://www.rand.org/well-being/social-and-behavioral-policy/centers/aging/dataprod/hrs-data.html

Helsing, K., Szklo, M., & Comstock, G. (1981). Factors associated with mortality after widowhood. *American Journal Of Public Health, 71(8),* 802-809. doi: 10.2105/ajph.71.8.802

Kaplan, G., Pamuk, E., Lynch, J., Cohen, R., & Balfour, J. (1996). Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ, 312(7037),* 999-1003. doi: 10.1136/bmj.312.7037.999

Klein, J., & Moeschberger, M. (2005). Survival analysis. New York: Springer.

Lee, E., Wei, L., Amato, D., & Leurgans, S. (1992). Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations. *Survival Analysis: State Of The Art*, 237-247. doi: 10.1007/978-94-015-7983-4_14

Li, J., Vestergaard, M., Cnattingius, S., Gissler, M., Bech, B., Obel, C., & Olsen, J. (2014). Mortality after Parental Death in Childhood: A Nationwide Cohort Study from Three Nordic Countries. *Plos Medicine, 11(7),* e1001679. doi: 10.1371/journal.pmed.1001679

Lin, D. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics In Medicine, 13(21),* 2233-2247. doi: 10.1002/sim.4780132105

Manor, O., & Eisenbach, Z. (2003). Mortality after spousal loss: are there socio-demographic differences?. *Social Science & Medicine, 56(2),* 405-413. doi: 10.1016/s0277-9536(02)00046-1

Martikainen, P., & Valkonen, T. (1996). Mortality after the death of a spouse: rates and causes of death in a large Finnish cohort. *American Journal Of Public Health*, 86(8_Pt_1), 1087-1093. doi: 10.2105/ajph.86.8_pt_1.1087

McGilchrist, C. (1993). REML Estimation for Survival Models with Frailty. *Biometrics, 49(1),* 221. doi: 10.2307/2532615

Oh, E., Shepherd, B., Lumley, T., & Shaw, P. (2017). Considerations for analysis of time-to-event outcomes measured with error: Bias and correction with SIMEX. *Statistics In Medicine, 37(8),* 1276-1289. doi: 10.1002/sim.7554

Pennebaker, J. (1985). Traumatic experience and psychosomatic disease: Exploring the roles of behavioural inhibition, obsession, and confiding. *Canadian Psychology/Psychologie Canadienne, 26(2),* 82-95. doi: 10.1037/h0080025

Rostila, M., & Saarela, J. (2011). Time Does Not Heal All Wounds: Mortality Following the Death of a Parent. *Journal Of Marriage And Family, 73(1),* 236-249. doi: 10.1111/j.1741-3737.2010.00801.x

Saarela, J., & Rostila, M. (2018). Mortality after the death of a parent in adulthood: a register-based comparison of two ethno-linguistic groups. *European Journal Of Public Health.* doi: 10.1093/eurpub/cky189

Smith, K., Hanson, H., Norton, M., Hollingshaus, M., & Mineau, G. (2014). Survival of offspring who experience early parental death: Early life conditions and later-life mortality. *Social Science & Medicine, 119,* 180-190. doi: 10.1016/j.socscimed.2013.11.054

Sonnega, A., Faul, J., Ofstedal, M., Langa, K., Phillips, J., & Weir, D. (2014). Cohort Profile: the Health and Retirement Study (HRS). *International Journal Of Epidemiology, 43(2),* 576-585. doi: 10.1093/ije/dyu067

Thiébaut, A., & Bénichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics In Medicine, 23(24),* 3803-3820. doi: 10.1002/sim.2098