Longitudinal analysis of mortality risk factors for actuarial valuation

Ushani Dias and Emiliano A. Valdez

Statistics Colloquium, Northern Illinois University, 26-27 April 2012

Outline

Introduction
- Global trends
- Literature
- Motivation

Data
- HRS survey
- Model construction
- Description

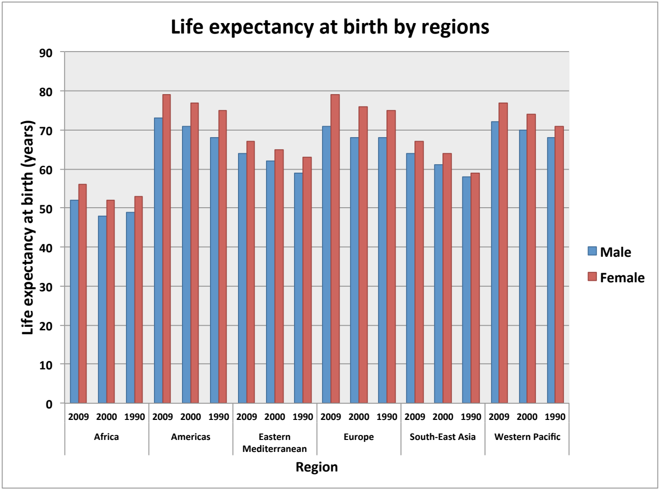
Survival models
- Model estimates
- Comparison

Additional work

Introduction

- There is no denying that the assumption of mortality plays a key role in the actuarial valuation of life insurance and annuity products.
- Within the last century alone, significant mortality improvement across several countries in the world have been due to:
  - significant medical progress
  - socio-demographic changes
  - improvements in lifestyles
  - the absence (or lack) of major pandemic crisis
- As a result, longevity poses a high risk to the insurance industry, something also that many involved in the industry have less understanding of its impact (economic or otherwise).

Global trends



Source: World Health Organization, 2012.

Literature - relevant publications

- Actuarial science: Kwon, H.-S. and B. Jones, 2005. "The Impact of the determinants of mortality on life insurance and annuities". *Insurance: Mathematics and Economics*, 38(2).
- Actuarial science: Fong, J. HY, 2010. "Beyond Age and Sex: Enhancing Annuity Pricing". *http://www.pensionresearchcouncil.org/publications/document.php*
- Medicine: Paula, M.L. et al., 2010. "Socioeconomic and behavioral risk factors for mortality in a national 19-year prospective study of U.S. adults". *Social Science & Medicine*, 70.
- Gerontology: Eileen, M. C. et al., 2010. "Mortality and morbidity trends: is there compression of morbidity?". *The Journal of Gerontology*, 66B.

Literature - continued

Useful books on modeling framework:

- Thomas R. Fleming, et al. (2005): Counting Processes and Survival Analysis
- Rogers R.G. et al. (2011): International Handbook for Adult Mortality

Relevant work

- International Actuarial Association (IAA) Mortality working Group
- "Global mortality improvement experience and projection techniques" by Purushotham et al. (2011), SOA sponsored research project.
- A survey work by Brown et al. (2003) with 45 recent papers provides some key factors that affect mortality after retirement.

Motivation

- In addition to age and sex, various studies have discovered significant effects of
  - demographic risk factors
  - health indicators
  - lifestyle factors
  - financial factors

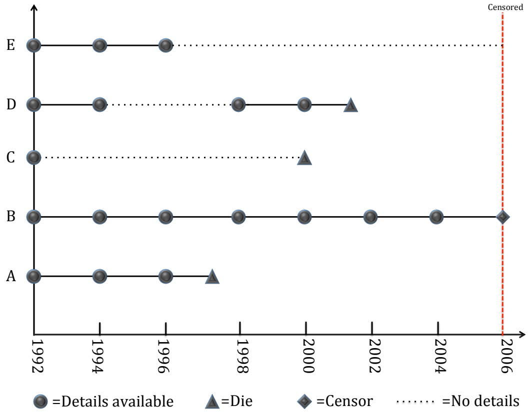
  on the mortality of both older and younger adults.
- We envision that the intention of our work is to:
  - identify (additional) significant risk factors affecting longevity
  - explore the association of significant covariates with survival distributions
  - understand how the various risk factors may possibly affect the values of annuity

Health And Retirement Study (HRS) Data

- HRS is a collaborative work between the University of Michigan, the National Institute of Aging, and the Social Security Administration.
- HRS is a prospective national longitudinal study about the health, retirement, and economic status of (some) Americans over the age 50 years.
- The study contains a rich amount of information that will allow us to explore both the cross-sectional and the longitudinal effects of various risk factors on mortality from 1992 to 2006.
- Awareness about the HRS data within the scientific community shows a rapid growth of its use in research.

Motivation for model construction

- Data-driven. Our observable is best illustrated by the following figure:



- This diagram provides an illustration of the nature of the HRS data.

Data description

- The HRS data is a survey from the general population.
- The data set contains 7,607 non-institutionalized financially responsible adults living in the contiguous United States in 1992.
  - follow-up studies were done every 2 years until 2006
- To better represent the U.S. population, sampling weights are used.
- Mortality data can be obtained from the National Death Index through 2006.
- Statistical analyses were conducted using SAS 9.3.

Demographic variables

Categorical Variables	Description		Proportions	
RAGENDER	Gender of the respondent:	Male=1	50.95%	
		Female=2	49.05%	
RARACEM	Race of the respondent:	White/Caucasian = 1	77.67%	
		Black /African American= 2	18.42%	
		Other= 3	3.91%	
RAEDUC	Education:	College and above= 0	39.46%	
		High-school graduate = 1	36.30%	
		Lt High-school = 2	24.24%	
RAVETRN	Veteran status:	No = 0	70.61%	
		Yes = 1	29.39%	
RMARRY	Current Partnership Status:	Single=0	33.41%	
		Married/Partnered=1	66.59%	
CENREG	Census Region:	Northeast = 1	16.90%	
		Midwest = 2	23.95%	
		South = 3	42.47%	
		West = 4	16.68%	
CENSOR	Censoring indicator for death:	Alive = 0	77.81%	
		Died= 1	22.19%	
Continuous variables		Minimum	Mean	Maximum
HKIDS	Number of living children of household	0	3.35	20
AGE	Age of the respondent	27	61	88
Date		Minimum	Mean	Maximum
RANYEAR	NDI death year	1992	1999	2006
RABYEAR	Birth year of the respondent	1912	1936	1965

Health variables

Categorical Health Variables	Description		Proportions
HBP	Reports high blood pressure:	No=0	51.35%
		Yes=1	48.65%
DIAB	Reports diabetes :	No=0	84.33%
		Yes=1	15.67%
CANCR	Reports cancer:	No=0	90.67%
		Yes=1	9.33%
LUNG	Reports lung disease:	No=0	90.30%
		Yes=1	9.70%
HEART	Reports heart problem:	No=0	82.42%
		Yes=1	17.58%
STROK	Reports stoke:	No=0	95.27%
		Yes=1	4.73%
PSYCH	Reports psychiatric problems :	No=0	85.34%
		Yes=1	14.66%
ARTHR	Reports arthritis problems :	No=0	47.70%
		Yes=1	52.30%

Lifestyle and Financial variables

Categorical Lifestyle Variables	Description		Proportions	
SMOKEV	Smoking Status	Non-smoker=0	35.80%	
		Former smoker=1	43.44%	
		Current smoker=2	20.75%	
DRINKR	Alcohol Drinking Status	<1 drink per day=0	50.40%	
		1-2 drinks per day =1	34.63%	
		≥ 3 drinks per day=2	5.97%	
VIGACT	Physical activity or Exercise 3+ times a week:	No=0	64.70%	
		Yes=1	35.30%	
Continuous Lifestyle Variable		**Minimum**	**Mean**	**Maximum**
BMI	Body Mass Index (kg/m^2)	10.80	27.75	102.70
Categorical Financial Variable	**Description**		**Proportions**	
JPHYS	Current job requires physical effort:	All the time=1	9.86%	
		Most of the time=2	8.78%	
		Some of the time=3	15.35%	
		None=4	18.83%	
		Does not apply=5	47.18%	
Continuous Financial Variables		**Minimum**	**Mean**	**Maximum**
HTOTW	Total Wealth(Excluding IRAs)	-4,733,000	252,167	85,960,000
HITOT	Total household income	0	51,619	7,395,294

Survival models

- Analyzes the time to event data.
- Applications in many different fields (e.g. Sociology, Engineering, Economics, Actuarial).
- Can be performed with retrospective or prospective data.
- Censoring and time-dependent covariates are two common features.
- Four general types of models:
  - Parametric (e.g. Gompertz , Weibull)
  - Nonparametric (e.g. Life table)
  - Semiparametric (e.g. Cox)
  - Discrete (e.g. Logit, Probit)
- For semiparametric models, martingale methods can be used.

Censored data regression models

Consider the right-censored failure time data for independent observations on $(X, \delta, \mathbf{Z})$ where

- $X = \min(T, U)$, T and U are failure and censoring times, respectively;
- $\delta = I_{[T \leq U]}$ indicator for failure; and
- $\mathbf{Z}$ is a p-dimensional column vector of covariates.

The information of

$$(X, \delta) \Rightarrow N(t) = I_{[X \leq t, \delta = 1]} \text{ and } Y(t) = I_{[X \geq t]}.$$

This setting leads to two possible approaches to censored regression models:

- traditional approach (Cox, 1972)
- counting process approach (Andersen et al.,1982)

The counting process approach

Consider the stochastic basis with the right continuous filtration $\{F_t : t \geqslant 0\}$ defined as

$$F_t = \sigma \{\boldsymbol{Z}, N(u), Y(u+) : 0 \leq u \leq t\}$$

- According to the Doob-Meyer Decomposition, for the increasing process N, there is a unique predictable process A with respect to F_t such that $N - A$ is a martingale.
- When $A^{'}$ exists, it is called the intensity process for N.
- Aalen (1978) shows that

$$\lim_{h \to 0} \frac{1}{h} \mathsf{Pr} [N(t+h) - N(t) = 1 | F_t] = \lambda(t+)$$

  where

$$\lambda_i(t) = Y_i(t)\lambda_0(t) \exp[\beta_0 \mathbf{Z_i(t)}]$$

The Andersen-Gill model

N has random intensity process λ such that

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp[\beta_0\mathbf{Z_i}(\mathbf{t})] = Y_i(t)\lambda\left\{t \mid \mathbf{Z_i}(\mathbf{t})\right\}$$

where

- $Y_i(t)$ is a predictable process taking values $\{0, 1\}$,
- λ_0 is a fixed underlying hazard function,
- β_0 is a fixed column vector of p coefficients, and
- $\mathbf{Z_i}$ is a column vector of p covariates.

Indeed, the Andersen-Gill model is a superset of the (familiar) Cox model.

Partial likelihood estimation technique

- To estimate β_0, partial (Cox's) likelihood techniques were employed. (Cox, 1975)
- Partial likelihood for n independent triplets $(N_i, Y_i, \mathbf{Z_i})$ where ties in observed failure times are allowed and for $i = 1, 2, \ldots, n$, we have

$$L(\beta, t) = \prod_{i=1}^{n} \prod_{s \geq 0} \left\{ \frac{Y_i(s) \exp[\beta' \mathbf{Z_i}(\mathbf{s})]}{\sum_{j=1}^{n} Y_i(s) \exp[\beta' \mathbf{Z_i}(\mathbf{s})]} \right\}^{\Delta N_i(s)}$$

  where

$$\Delta N_i(s) = 1, \text{if } N_i(s) - N_i(s-) = 1,$$

  and otherwise, $\Delta N_i(s) = 0$.
- Andersen et al. (1982) and Fleming et al. (2005)

Model estimates - based on the likelihood technique

Variable	Parameter Estimate	Standard Error	Pr>ChiSq	Hazard Ratio
RAGENDER- Male	0.56318	0.08106	<.0001	1.756
RMARRY- Single	0.20008	0.08112	0.0136	1.222
AGE	0.04323	0.00744	<.0001	1.044
AGE-Unit 5				1.241
DIAB- Yes	0.75472	0.07885	<.0001	2.127
LUNG- Yes	0.46491	0.08532	<.0001	1.592
HEART- Yes	0.40177	0.07715	<.0001	1.494
STROK- Yes	0.58143	0.09791	<.0001	1.789
CANR- Yes	0.95014	0.08067	<.0001	2.586
VIGACT- No	0.82516	0.09596	<.0001	2.282
DRINKR- Mod	-0.36612	0.09005	<.0001	0.693
DRINKR- Heavy	-0.31438	0.14643	0.0318	0.730
SMOKEV- Former	0.41537	0.09304	<.0001	1.515
SMOKEV- Current	0.66674	0.10522	<.0001	1.948
BMI	-0.05391	0.00732	<.0001	0.948
BMI-Unit 5				0.764
JPHYS-Most	-0.10610	0.23708	0.6545	0.899
JPHYS-Some	-0.14962	0.20673	0.4692	0.861
JPHYS-None	-0.30787	0.20670	0.1364	0.735
JPHYS-NA	0.53836	0.17051	0.0016	1.713
HITOT	-3.7916E-6	1.10751E-6	0.0006	1.000
HITOT-Unit 50000				0.827

Variable selection results - comparison

Demographic variables	Agree or not	Literature
AGE	√	Horuchi S. et al.,2010; Brown R.L., 1988
RAGENDER	√	Rogers R.G., 1995; Travato, F., & N. K. Lalu, 1998
RAEDUC	√ ×	Paula M.L. et al.,2010; Sorlie P.D. et al., 1995 Attanasio O.P., & and C. Emmerson, 2001
RARACEM	√ ×	Kallan J ., 1997; Attanasio O.P., & and C. Emmerson, 2001 Williams D.R.& C. Collins, 1995;Hummer R.A., 1996
RAVETRN	√	Alex H.S.H., & C.E. Thoresen, 2005
RMARRY	√ ×	Hui Liu, 2009 ; Kaplan R.M., & Richard H.K., 2006 Attanasio O.P., & and C. Emmerson, 2001; Rogers R.G.,1995
CENREG	√	Purushotham M., et al.,2011
HKIDS	√	Kotler P., & D.L.Wingard, 1989

Health varialbes	Agree or not	Literature
HBP	√	Gu Q. et al., 2007; National Vital Statistics Report, 2009
DIAB	√	Shaista M. et al., 2004; National Vital Statistics Report, 2009
LUNG	√	Mannino D.M., 2003; National Vital Statistics Report, 2009
HEART	√	Shaista M. et al., 2004; National Vital Statistics Report, 2009
STROK	√ ×	National Vital Statistics Report, 2009 Joelle HY. Fong, 2010
PSYCH	×	Joelle HY. Fong, 2010;
CANR	√	National Vital Statistics Report, 2009
ARTHR	√ ×	Kroot E.J.A. et al., 2000 Doran M.F. et al., 2002; Avina Zubieta J.A. et al., 2008

Variable selection results - comparison

Lifestyle variables	Agree or not	Literature
VIGACT	$\surd$	Doll R. et al., 2004; Steven N.B., 1996
DRINKR	$\surd$ $\times$	Thun M.J. et al., 1997; Paula M.L. et al.,2010 Valliant G.E., & K.Mukamal, 2001
BMI	$\surd$ $\times$	Campos et al., 2006; Sui et al., 2007 Wei et al., 1999;
SMOKEV	$\surd$	Doll R. et al., 2004; Lantz et al., 1998

Financial variables	Agree or not	Literature
JPHYS	$\surd$ $\times$	Valliant G.E., & K. Mukamal, 2001 Brown R.L., 1997
HTOTW	$\times$	Attanasio O.P. et al., 2000; Menchik Paul 1993
HITOT	$\surd$ $\times$	Moulton B.E. et al., 2012; Krieger N. et al., 2005 Blakely T. et al., 2003

Future work

- Enhance the variable selection process (e.g. Bayesian variable selection)
- Fit alternative parametric survival models for comparison purposes
- Incorporate missing data imputation methods
- Examination of financial or economic impact:
    - the possibility of natural hedging between life insurance and life annuity products
    - other insurance products such as long term care