Analysis on Mortality Cohort Effect of Birth Year in view of Differential Geometry and its Application

Ning Zhang[1] Liang Zhao [2]

[1]China Institute for Actuarial Science
Central University of Finance and Economics

[2]School of Mathematical Sciences
Beijing Normal University

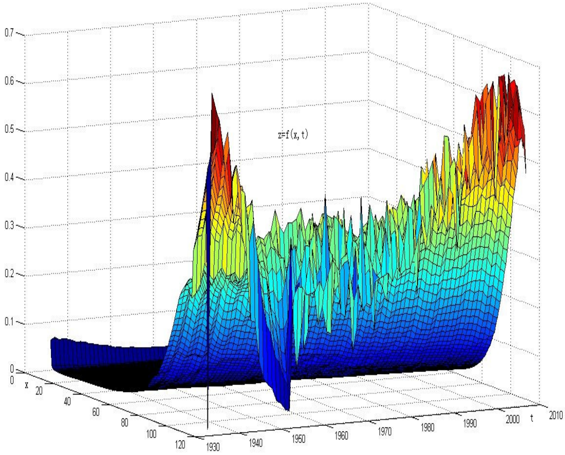
September 7, 2015

Introduction

What can we do ?

- Input: Mortality data sets, (Mortality Table, age-specific and period-specific)
- We can find which one in different generations has most different character in mortality. Or how different is it from the others.
- Or which one has experienced more obvious mortality cohort effect in several mortality data sets.
- Or maybe it is beneficial to decide whether or not we consider cohort effect into mortality model.

An example of mortality surface of U.S.



Background

- Background 1, mortality models : Mortality cohort effect of birth year has attracted widespread attentions.

  it is well known that people born in the U.K. between 1925 and 1945 have experienced more rapid improvement in mortality than generations born in other periods . In other words, this generation has experienced stronger cohort effect than others. [2, 3]

- Background 2, data visualization: Detection and measurement directly from data sets is desired in demography or other fields.

  Mortality cohort effect of birth year, different birth months, or different cities

- Background 3, Big Data: Information discovery or Data Mining based on data sets becomes widespread.

  Generations with special characters, birth year with special character, age with special character and so on

Motivation

- We want to directly detect the generation whose mortality is different from others.
- From Data sets to Surface, Mortality difference was showed on surface by vision.
- Curvature, normal vector etc were linked to this task: differential geometry.
- An intuitive example:

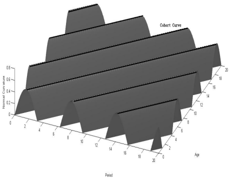


Figure: The sample of cohort curve which is not real, just for

Definition

Definition: *For a cohort curve on Σ_s, suppose $T(s)$ is the tangent vector of a point $p(s) \in l_t$ and $N(s)$ is the orthogonal direction of $T(s)$ on the tangent plane of $p(s)$. Let $NC_T(s)$ and $NC_N(s)$ be the normal curvatures along the directions $T(s)$ and $N(s)$ respectively. Then the integral*

$$CEI_t = \int_a^b |NC_T(s) - NC_N(s)| ds \tag{2.1}$$

is called the cohort effect index(CEI) of the generation born in year t. Here, for simplicity, we use the arc length s as parameter to describe the cohort curve l_t and the integrating range $[a, b]$ is decided by the structure of the mortality data.

Computing CEI on the discrete surface (or mortality data sets)

And the following will compute the normal curvature on the discrete surface (Mortality surface);

- 1: Discrete tangent vector (Evaluating tangent vector on discrete point)
- 2: Discrete curvature vector (Evaluating curvature vector on discrete point)
- 3: Discrete normal vector (Evaluating normal vector on discrete point)
- 4: Discrete normal curvature (Evaluating normal curvature on discrete point)
- 5: Get CEI

1&2, Discrete tangent vector and curvature vector

To begin our program, first we define the discrete parameter of a discrete curve l contains three point p_0, p_1 and p_2. We set

$$s_0 = 0, \quad s_1 = \frac{|p_1 - p_0|}{|p_1 - p_0| + |p_2 - p_1|}, \quad s_2 = 1.$$

Next we estimate the tangent vector of l at p_1. We call it a discrete tangent vector and denote it by $\vec{T} = (T_t, T_x, T_z)$.
By minimizing the sum of the distances between the tangent line and the two points p_0 and p_1 under the constrain that the tangent line should pass through the point p_1, we can get an approximation of $\vec{T}$.

1&2, Discrete tangent vector and curvature vector

$$T_t = \frac{(s_0 - s_1)(t(s_0) - t(s_1)) + (s_2 - s_1)(t(s_2) - t(s_1))}{(s_0 - s_1)^2 + (s_2 - s_1)^2}, \quad (2.2)$$

$$T_x = \frac{(s_0 - s_1)(x(s_0) - x(s_1)) + (s_2 - s_1)(x(s_2) - x(s_1))}{(s_0 - s_1)^2 + (s_2 - s_1)^2}, \quad (2.3)$$

$$T_z = \frac{(s_0 - s_1)(z(s_0) - z(s_1)) + (s_2 - s_1)(z(s_2) - z(s_1))}{(s_0 - s_1)^2 + (s_2 - s_1)^2}. \quad (2.4)$$

By theories of differential geometry, for a smooth curve l parameterized by s, suppose the unit tangent vector field along $l(s)$ is $\vec{V}(s)$, then the curvature vector of the curve is defined by

$$C\vec{V}(s) = \frac{\vec{V}'(s)}{|l'(s)|}, \quad (2.5)$$

1&2, Discrete tangent vector and curvature vector

For the two discrete curve l_1 and l_2, by formulas (2.2-2.4) and normalization, we can get the unit discrete tangent vectors to l_1 and l_2 at point p_{ij} and we denote them by
$\vec{V}_1(p_{ij}) = (v_{1t}(p_{ij}), v_{1x}(p_{ij}), v_{1z}(p_{ij}))$ and
$\vec{V}_2(p_{ij}) = (v_{2t}(p_{ij}), v_{2x}(p_{ij}), v_{2z}(p_{ij}))$.
For a discrete curve, the derivative with respect to its discrete parameter can be defined by solving a similar constrained minimization problem as we do in estimating $\vec{T}$. Thus we can get the two discrete curvature vector fields $\vec{CV}_1$ and $\vec{CV}_2$ just following the formula (2.5).

Thirdly, Discrete normal vector

Obviously, two unit tangent vectors $\vec{V}_1$ and $\vec{V}_2$ are not enough to determine a unique vector orthogonal to them. To get the normal vector of the surface Σ_d at any point p_{ij}, we consider two more short discrete curves across p_{ij}.

Let $l_3 : \{p_{i-1,j}, p_{ij}, p_{i+1,j}\}$ and $l_4 : \{p_{i,j+1}, p_{ij}, p_{i+1,j-1}\}$, the same as we do for l_1 and l_2, we can get two unit tangent vectors $\vec{V}_3$ and $\vec{V}_4$.

Since normal vector are orthogonal to any tangent vector, we can estimate the discrete unit normal vector $\vec{N}(p_{ij})$ by minimizing

$$f(\vec{N}) = \sum_{k=1}^{4} |\vec{N} \cdot \vec{V}_k|^2,$$

with the constraint $\vec{N} \cdot \vec{N} = 1$.

Fouthly, Discrete normal curvature

It is nature to define the discrete normal curvature along direction $\vec{V}_k$ at point p_{ij} by

$$NC_k(p_{ij}) = N(p_{ij}) \cdot \vec{CV}_k(p_{ij}), \quad k = 1, 2, 3, 4$$

Finally, CEI

For a fixed integer m, all the points p_{ij} satisfying $i + j = m$ make up a curve related to persons born in the same year. We call these persons cohort m,or C_m and call the curve a cohort curve.
The tangent vector field along the cohort curve corresponds to $\vec{V}_1$ and we call this direction a cohort direction. By our definition of CEI for the smooth case, we define the discrete CEI for C_m by

$$CEI_m = \sum_{i+j=m} |NC_1(p_{ij}) - NC_2(p_{ij})|.$$

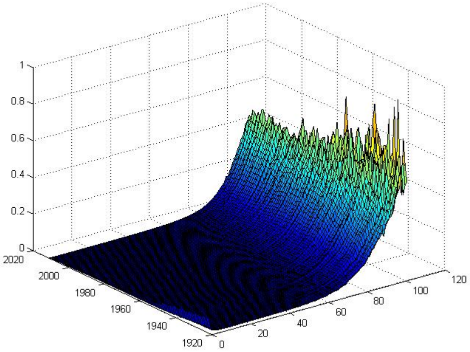
Series of Cohort Effect: the time series including the cohort effect indexes of all the birth years (Generations)

Data

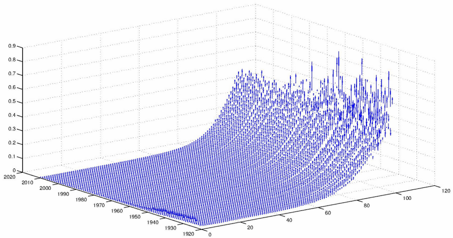
Our data come from the "The Human Mortality Database" [1]. Several types of data sets can be used. We choose the data sets of "Death rate" and "1 × 1" for our practice and we also use the data sets of "1 × 5" and "5 × 5" for auxiliary check or comparison. The following four figures showed the computing process of CEI (U.K.)

[1] http://www.mortality.org/

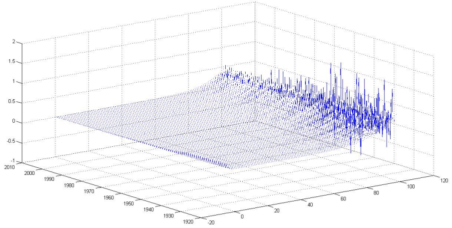
U.K.: Mortality surface



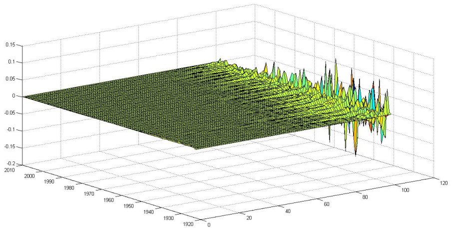
U.K.: Tangent vector field (at one direction)



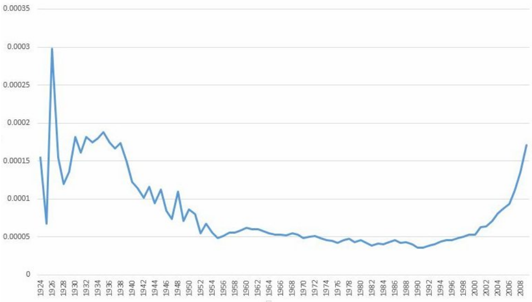
U.K.: Curvature vector field (at one direction)



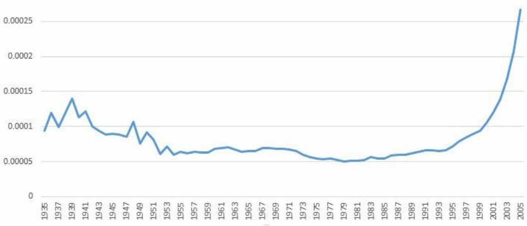
U.K.: Normal curvature



U.K. Cohort Effect Indexes, Series of Cohort Effect

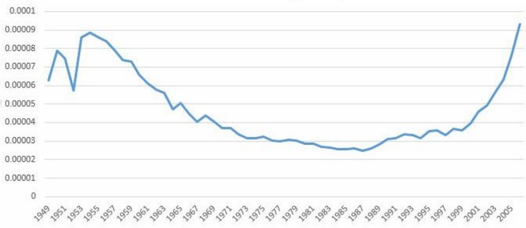


U.S. Cohort Effect Indexes, Series of Cohort Effect

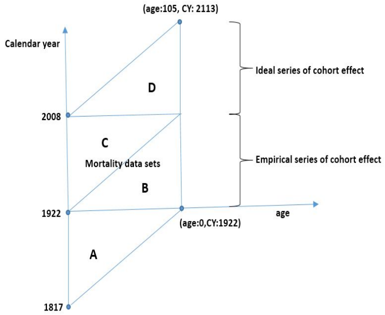


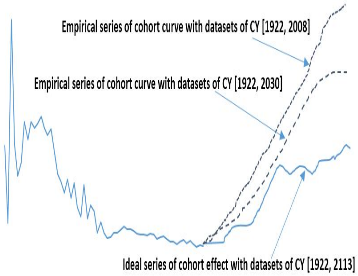
Canada, Cohort Effect Indexes, Series of Cohort Effect

Japan, Cohort Effect Indexes, Series of Cohort Effect



Ideal and Empirical Series of Cohort Effect





Ideal series of cohort effect: We can get an ideal series of cohort effect by using all need data which includes Part A, Part B, Part C and Part D.

Empirical series of cohort effect: the usual mortality data set is a data matrix which consists of Part B and Part C.

Aggregating Index of Cohort Effect

Definition: For any empirical series of cohort effect, the aggregating index of cohort effect is defined as its coefficient of variation, in other words, the quotient of its sample standard deviation divided by its sample mean.

Table: Aggregation index of cohort effect

	US	Canada	Japan	UK
mean	8.14646E-05	8.31718E-05	4.65899E-05	8.24972E-05
Variance	1.32031E-09	2.31223E-09	4.04465E-10	2.59693E-09
stdev	3.63361E-05	4.80857E-05	2.01113E-05	5.09601E-05
CV	0.446035861	0.578148277	0.431667221	0.617719261

Generation Gap

Another derivative parameter from series of cohort effect is the generation gap (or cohort-effect generation gap) which describes how long cohort effect maintains. Formally, it is the gap from the beginning to the end of a peak on the series of cohort effect. So there are many generation gaps in view of cohort effect on the ideal series of cohort effect. And we recommend to consider only the part before 1970 when using empirical series of cohort effect.
Usually, in social sciences, we use 5 or 10 years to represent a generation. But our results show that the length of a generation is a problem in itself and we give a method to resolve the problem. Table 2 gives the maximal and minimal generation gaps for the four targeted countries. This is also an important character for the population.

Applications of our model in the analysis on longevity risk are one of the problems we are considering. Since the series of cohort effect measure the strength of cohort effect for different generations, we can introduce a parameter into classical mortality models: for example, Lee-Carter model may be changed into $\ln \mu_x(t) = \alpha_x + \beta_x k_t + \sigma c_{i-x}$, where c_{i-x} is decided by the series of cohort effect.

Conclusion

We promote an effective method based on differential geometry to implement quantitative measurement of cohort effect. The peaks on the series of cohort effect mean the existence of cohort effects and the height of the peaks tells us the strength of the corresponding effects.

We also apply this method on the data sets of four countries including the United Kingdom, the United states, Canada and Japan. All the resulting series show the desired strength of cohort effects in different generations.

In particular for U.K., our method can give a further description of the well-known mortality cohort effect. Based on the series of cohort effect, we introduced the aggregating index of cohort effect (AICE) which is a general description of cohort effect for the whole population of a country or group.

Thanks for your attention

Welcome any advice or questions to Email:
12057866@qq.com
Or
nzhang@amss.ac.cn

Continous mortality investigation (2002). Working paper 1: An interim basis for adjusting the mortality projections for cohort effects. C.M.I.B., London.

K. Navaneetham (1993). The Influence of Cohort Effects on Mortality Trends in India: Role of Economic Factors, Population Research and Policy Review,12(2), 159-176

Willets, R. C. (2004). The cohort effect: insights and explanations. British Actuarial Journal, 10, 833-877.

Lee, R. D. and Carter, L. R. (1992). Modelling and forecasting U.S.mortality. Journal of the American Statistical Association, 87(14), 659-675.

Lee, R. D. (2000). The LeeCCarter method for forecasting mortality, with various extensions and applications. North American Actuarial Journal, 4(1), 80-93.

Brouhns, N. and Denuit, M. (2002). Risque de longvit et rentes viagres. II. Tables de mortalit prospectives pour la population belge. Belgian Actuarial Bulletin, 2, 49-63.

Brouhns, N., Denuit, M., and Keilegom, van, I. (2005). Bootstrapping the Poisson log-bilinear model for mortality forecasting. Scandinavian Actuarial Journal, (3), 212-224.

Renshaw, A. E. and Haberman, S. (2008). On simulation-based approaches to risk measurement in mortality with specific reference to poisson Lee C Carter modelling. Insurance: Mathematics and Economics,42, 797-816.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee - Carter model for mortality reduction factors. Insurance:Mathematics and Economics, 38(3), 556-570.

Winship, C., Harding, D. (2008). A mechanism based approach to the identification of Age C Period C Cohort models. Sociological Methods and Research 36, 362-401.

National Statistics-Government Actuarys Department (2001). National population projections: Review of methodology for projecting mortality. National Statistics Quality Review Series, Report No. 8.

Olivieri, A. (2001). Uncertainty in mortality projections: an actuarial perspective. Insurance: Mathematics and Economics, 29(2), 231-245.

Pollard, J. H. (1987). Projection of age-specific mortality rates. Population Bulletin of the UN, 21-22, 55-69.

Guillot, M. (2003). The cross-sectional average length of life (CAL): A crosssectional mortality measure that reflects the experience of cohorts. Population Studies, 57(1), 41-54.

Schoen, R., Canudas-Romo, V. (2005). Changing mortality and average cohort life expectancy. Demographic Research, 13(5), 117-142.

N. J. Mitra, A. Nguyen, Estimating surface normals in noisy point cloud data, in: Proceedings of the Nineteenth Annual Symposium on Computational Geometry, 2003, 322-328