WHY DOES A HUMAN DIE? A STRUCTURAL APPROACH TO COHORT-WISE MORTALITY PREDICTION UNDER SURVIVAL ENERGY HYPOTHESIS

BY

YASUTAKA SHIMIZU, YUKI MINAMI AND RYUNOSUKE ITO

ABSTRACT

We propose a new approach to mortality prediction under *survival energy hypothesis (SEH)*. We assume that a human is born with initial energy, which changes stochastically in time and the human dies when the energy vanishes. Then, the time of death is represented by the first hitting time of the survival energy (SE) process to zero. This study assumes that SE follows a time-inhomogeneous diffusion process and defines the mortality function, which is the first hitting time distribution function of the SE process. Although SEH is a fictitious construct, we illustrate that this assumption has the potential to yield a good parametric family of cumulative probability of death, and the parametric family yields surprisingly good predictions for future mortality rates.

KEYWORDS

Mortality prediction, survival energy hypothesis, the first passage time, diffusion processes, ruin theory.

JEL codes: 62M86, 91B30, 60G25.

1. INTRODUCTION: WHY DOES A HUMAN DIE?

Mortality prediction is an important social problem. The most popular mortality prediction model is the Lee–Carter model (Lee and Carter, 1992), which is based on death rates. Let τ_x be the time of death of an individual of age x. Traditionally, τ_x is supposed to be the first event of an inhomogeneous Poisson process with the force of mortality $\mu(t, x)$ at time t. Let $m(x, t)$ be the *crude death rate* of age x and let the calendar year be t. Then, assuming $\mu(x, t)$

remains constant over each year of integer age x and over each calendar year t, we see that $\mu(t, x) = m(t, x)$, and the *mortality rate* is given by

$$
\begin{aligned}
{}_{t|}q_x &:= \mathbb{P}(\tau_x \leq t+1 \,|\, \tau_x > t) \\
&= 1 - \exp\left(-\int_t^{t+1} \mu(s, x)\, ds\right) = 1 - \exp(-m(x, t)).
\end{aligned}
$$

The Lee–Carter model is a regression type model of this mortality rate:

$$
\log m(t, x) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{1.1}
$$

where α_x, β_x are time-independent parameters, $\kappa = (\kappa_t)_{t \geq 0}$ is a latent time series, and $\epsilon_{x,t}$ is an i.i.d. noise. There are many other extensions or modifications of this model. For example, Renshaw and Haberman (2003) considered the multifactor version of the Lee–Carter model, Renshaw and Haberman (2006) and Cairns *et al.* (2009) considered cohort effects, and Chen and Cox (2009) include jump effects. See the review by Cairns *et al.* (2008) for other studies.

    Those models concentrate on "probability" by considering death an accidental event, but they do not explain why a human dies. An analogy can be seen in credit risk analysis, where τ is considered the time of default of a company. Such a model is called the *reduced-form approach* in credit risk analysis; see, for example, Jarrow and Turnbull (1995). Except for the Lee–Carter type approaches, there are other reduced-form approaches in mortality prediction, as in Dahl (2004), Biffis (2005), Biffis *et al.* (2010), Bauer *et al.* (2012) (see also the references therein).

    On the other hand, there is a *structural approach*, where the "firm value" of the company is modeled by a stochastic process, and where the time of default is defined as the first passage time of the process to a certain region (the *Merton model*), which is initiated by Merton (1974, 1976). There are many extensions to this stream; see, for example, Schoutens and Cariboni (2009) and references therein. This is also the same as the probability of ruin in insurance risk theory, which is originally introduced by Lundberg (1903), where the ruin occurs if an insurance surplus reaches the negative region. Moreover, there is also a similar idea in the analysis of a system failure in engineering, where the "failure time τ" appears if accumulating "damages" exceed a certain level; for example, Ye and Chen (2014), among others.

    In our study, we will use the same idea as those "structural approach" in mortality forecasting. That is, we assume that there is an underlying *survival energy (SE)* for human beings, and that a human dies when that energy vanishes: survival energy hypothesis (SEH); see also Ito and Shimizu (2019). The flow of energy would depend on the *cohort* in which a human was born because it would depend on the period: natural, economic, and social environments, technology and medical care levels, and so on. Therefore, we will consider cohort-wise SE dynamics.

Mathematically, we suppose that SE at time t in a cohort c is given by a random variable X_t^c with initial energy $X_0^c = x_c$. The sample space Ω is a family in the corresponding cohort, and each sample $\omega \in \Omega$ corresponds to a person. That is, for a person ω, his/her energy has a different flow $X_\cdot^c(\omega)$, but the law is governed by an underlying probability law $\mathbb{P}$ on the σ-field $\mathcal{F}$ of Ω. In this way, $(\Omega, \mathcal{F}, \mathbb{P})$ becomes a probability space and $X^c = (X_t^c)_{t \geq 0}$ is a stochastic SE process in that space.

Later, we will assume that the SE process follows a stochastic differential equation with an unknown parameter, which means that every human in the fixed cohort follows the same SE model but has a different path due to randomness. Therefore, we cannot separate the individual SE processes, but we can compare the mortality changes in one cohort with that of the others.

In the sequel, we assume a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{F})$ with the usual conditions, where $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, is given, and that the cohort-wise SE process $X^c = (X_t^c)_{t \geq 0}$ is a $\mathbb{F}$-adapted process. The time of death of a human in the cohort c is given by the first hitting time of X^c to zero:

$$\tau^c = \inf\{t > 0 : X_t^c < 0\}, \tag{1.2}$$

which can be the $\mathbb{F}$-stopping time under some regularities and the probability of death up to time t is given by

$$q_c(t) = \mathbb{P}(\tau^c \leq t), \quad t > 0. \tag{1.3}$$

We call it the *mortality function*.

Our setting for the mortality function is similar to those for the default and ruin probabilities. That is, our analysis is essentially the one for the first hitting time to zero of the SE process X^c which is the basic structure for human beings. Therefore, we call this approach a *(cohort-wise) structural approach* to mortality prediction after the earlier literatures.

There are many practical advantages of identifying cohort-wise mortality function (1.3). For example, when we compute life expectancy at age x, say e_x, in practice,

$$\begin{aligned}
e_x &= p_x + {}_2p_x + {}_3p_x + \ldots \\
&= \sum_{t=1}^{\infty} {}_{t|}q_x + \sum_{t=2}^{\infty} {}_{t|}q_x + \sum_{t=3}^{\infty} {}_{t|}q_x + \ldots \\
&= \sum_{n=1}^{\infty} \sum_{t=n}^{\infty} [1 - \exp(-m(x, t))]
\end{aligned}$$

where ${}_tp_x$ is the actuarial notation meaning the survival probability of age x for t years. Since most conventional approaches for estimating $m(x, t)$ give a graph in x for a fixed t, we would need many graphs of $m(x, t)$ $(t = 1, 2, \ldots)$ to calculate e_x.

However, if we can obtain $q_c(t)$ in (1.3) as a closed analytical function of t, we can obtain e_x with a simple (numerical) integral

$$e_x = \int_x^\infty \frac{1 - q_c(t)}{1 - q_c(x)} \, \mathrm{d}t.$$

Moreover, the single premium of an entire life insurance policy at age x (with $\delta > 0$ interest force) is given by

$$\overline{A}_x = \int_0^\infty e^{-\delta t} \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{P}(x < \tau^c \le x + t | \tau^c > x) \, \mathrm{d}t$$

$$= \int_0^\infty e^{-\delta t} \frac{q_c'(t + x)}{1 - q_c(x)} \, \mathrm{d}t,$$

among others. Only a mortality function can give many cohort-wise actuarial quantities in simple computations.

The rest of this paper is organized as follows. Section 2 explains the structural approach for computing cohort-wise probability of death using a Brownian survival energy model (BSEM) for a simple understanding. We extend the toy model to a more general time-inhomogeneous diffusion process in a natural way and propose a specific model that can give the explicit form of the hitting time distribution function, which is called the *mortality function*. Section 3 gives a method for estimating the model parameters under restricted data situations that would be unique in our context. Section 4 is devoted to real-life data analysis, where cohort-wise mortality functions are estimated and predicted based on open data from the Human Mortality Database (https://www.mortality.org/). Lastly, we conclude with Section 5.

2. SEM: SURVIVAL ENERGY MODELS

2.1. Describing the idea with a toy SEM

To understand the structural approach that we use, let us consider a simple example of an SE model (SEM).

For simplicity, we assume that the SE of cohort c is a drifted Brownian motion:

$$X_t = x_c + \mu_c t + \sigma_c W_t, \quad t \ge 0, \tag{2.1}$$

where W is a Wiener process and μ_c, σ_c, and x_c are unknown parameters. Although this model is too simple for use in practice, it is useful to understand some of the advantages of the structural approach, which will be described in the sequel.

In this model, it is known that the hitting time distribution function (1.3) is given explicitly as follows:

$$\psi_t^\theta(x_c) = 1 - \Phi\left(\frac{x_c + \mu_c t}{\sigma_c \sqrt{t}}\right) + e^{-2\mu_c x_c/\sigma_c^2}\Phi\left(\frac{-x_c + \mu_c t}{\sigma_c \sqrt{t}}\right), \quad t > 0, \qquad (2.2)$$

where $\theta = (\mu_c, \sigma_c)$ and Φ are standard normal distribution functions; see Baukai (1990). If $\mu_c < 0$, the distribution of τ^c is an *inverse-Gaussian distribution* with mean $\mathbb{E}[\tau^c] = x_c/|\mu_c|$ and variance $x_c\sigma_c^2/|\mu_c|^3$. Otherwise, if $\mu_c \geq 0$, it holds that $\mathbb{E}[\tau^c] = \infty$.

2.1.1. *The meaning of the parameters*
- *Initial SE*: x_c is the initial SE of cohort c, which is common within the cohort.
- *SE drift*: μ_c is the average of increasing rate of SE since $\mathbb{E}[X_t] = x_c + \mu_c t$. Then, the life expectancy of that cohort is given by $\mathbb{E}[\tau^c] = x_c/|\mu_c|$.
- *SE volatility*: σ_c^2 can be a measure of a sudden death since the higher the σ_c^2, the easier it is for the energy process to hit zero.

2.1.2. *Interpretation of the model*
By estimating parameters (μ_c, σ_c^2, x_c) and observing those changes year on year for a fixed cohort, we may be able to interpret *"Why does a human die?"*

- If the initial SE x_c is estimated to be high, then it indicates that the corresponding cohort has a potential of longevity.
- If the estimated μ_c is increasing year on year, then it indicates that a human in that cohort will not die year after year. If it is estimated as $\mu_c \leq 0$, then it implies that $\mathbb{P}(\tau^c < \infty) = 1$, which will be usual. However, it would be interesting if it is estimated to be $\mu_c > 0$, which means that $\mathbb{E}[\tau^c] = \infty$: *a human will live forever!* Such a belief does exist in some areas of biological philosophy, and if it were to be true, we might be able to "prove" the belief statistically.
- If σ_c^2 is estimated to be high, it implies that there is a probability of a "sudden death" because the SE process easily hits zero. If the estimated σ_c^2 is decreasing year on year, then it indicates that a sudden death is going to decrease, which might also indicate an improved quality of life (medical care and lifestyle).

2.2. A practical point of view

It would not be realistic to assume that the parameters μ_c and σ_c^2 are constant throughout life as in the previous model. Intuitively speaking, in the first stage of life, childhood and teenage years, death rate and survival rate are unstable due to high volatility. In the middle stage, 20–40 years of age, μ_c may not be that high though σ_c^2 is somewhat high. In the last stage after retirement, μ_c
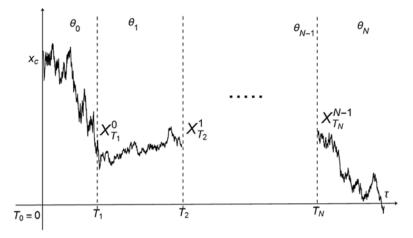
FIGURE 1: An image of the survival energy (SE) process.

and σ_c^2 decrease simultaneously; see Figure 1. Actually, the curve (2.2) cannot follow the empirical mortality function, which increases significantly in the last stage. Considering this, we shall consider a stochastic model where the drift and volatility depend on time.

2.3. Inhomogeneous diffusion SEM: Id-SEM

Hereafter, we assign a certain cohort as c and omit the index c.

On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0})$ with usual conditions, we assume that the SE process of a certain cohort, say $X^c = (X_t^c)_{t \geq 0}$ satisfies a stochastic differential equation

$$\mathrm{d}X_t^c = U(t, \mu_c)\, \mathrm{d}t + V(t, \sigma_c)\, \mathrm{d}W_t, \quad X_0^c = x_c, \tag{2.3}$$

where W is a Wiener process, $U : [0, \infty) \times \Xi$, $V : [0, \infty) \times \Pi$ are measurable functions, with $\Xi \subset \mathbb{R}^p$, $\Pi \subset \mathbb{R}^q$ for $p, q \in \mathbb{N}$ and we set $\theta = (x_c, \mu_c, \sigma_c) \in \Theta$ with $\Theta = \Lambda \times \Xi \times \Pi$, where $\Lambda \subset (0, \infty)$.

We will be using such a time-inhomogeneous diffusion (Id) process for the SE model (SEM) in this study, and we shall call this model *Id-SEM*. More specifically, we shall consider some *change point models*, where the parameters in U and V have several change points in time.

2.3.1. *Brownian SEM with 1 change point (1-BSEM)*
Ito and Shimizu (2019) consider the following SEM, say, $X^{c,1} = (X_t^{c,1})_{t \geq 0}$, which satisfies the SDE (2.3) with

$$U(t, \mu_c) = \mu_{c,0} \mathbf{1}_{(0,T]}(t) + \mu_{c,1} \mathbf{1}_{(T,\infty)}(t),$$
$$V(t, \sigma_c) = \sigma_{c,0} \mathbf{1}_{(0,T]}(t) + \sigma_{c,1} \mathbf{1}_{(T,\infty)}(t),$$

where $\mu_c = (\mu_{c,0}, \mu_{c,1})$ and $\sigma_c = (\sigma_{0,c}, \sigma_{c,1})$. This model is equivalent to the following piecewise drifted Brownian motion:

$$X_t^{c,1} = \begin{cases} X_t^{c,1} := x + \mu_{c,0}t + \sigma_{c,0}W_t, & t \in [0, T] \\ X_T^{c,0} + \mu_{c,1}(t - T) + \sigma_{c,1}(W_t - W_T), & t \in (T, \infty) \end{cases}.$$

That is, the process X has a change point T, and it has parameters $\theta_0 = (\mu_{c,0}, \sigma_{c,0})$, $\theta_1 = (\mu_{c,1}, \sigma_{c,1})$ and the initial value x.

In particular, this path is continuous since $X_{T_1}^1 = X_{T_1}^0$. This mortality function is the same as the hitting time distribution function computed in Baukai (1990).

2.3.2. *Brownian SEM with N change points*
Consider a generalization of 1-BSEM into the N-change point Brownian SEM (N-BSEM).

For a sequence $0 = T_0 < T_1 < \cdots < T_N < T_{N+1} = \infty$ for $N \in \mathbb{N}$, let

$$\begin{cases} U(t, \mu_c) = \displaystyle\sum_{k=0}^{N-1} \mu_{c,k}\mathbf{1}_{(T_k, T_{k+1}]}(t); \\[2mm] V(t, \sigma_c) = \displaystyle\sum_{k=0}^{N-1} \sigma_{c,k}\mathbf{1}_{(T_k, T_{k+1}]}(t), \end{cases} \tag{2.4}$$

where $\mu_c = (\mu_{c,k})_{k=0,\dots,N}$ and $\sigma_c = (\sigma_{c,k})_{k=0,\dots,N}$, and consider that $(T_N, T_{N+1}] = (T_N, \infty)$. Then the corresponding SE process, say, $X^{c,N}$, follows a piecewise (drifted) Brownian motion.

This is equivalent to $X^N = (X_t^N)_{t \geq 0}$ defined as the following recurrence system of stochastic equations: for $k = 1, 2, \dots, N$,

$$X_t^0 = x + \mu_{c,0}t + \sigma_{c,0}W_t, \quad t \in [0, T_1]; \quad X_t^0 \equiv X_{T_1}^0, \quad t > T_1;$$

$$X_t^k = \begin{cases} X_t^{k-1} & t \leq T_k; \\ X_{T_k}^{k-1} + \mu_{c,k}(t - T_k) + \sigma_{c,k}(W_t - W_{T_k}), & t \in (T_k, T_{k+1}]; \\ X_{T_{k+1}}^k, & t > T_{k+1} \text{ (if } k < N). \end{cases} \tag{2.5}$$

The mortality function of N-BSEM is given by a recurrence-type formula for the mortality functions for the process X^k's $(k = 1, 2, \dots, N)$; see Ito and Shimizu (2019) but it is difficult to implement in practice since the mortality function is not explicit. In reality, 1- and 2-BSEMs are fitted to real-life mortality data, and an *ad hoc* modification is used for the mortality function due to theoretical and practical difficulties. To overcome these difficulties, we shall use a model as a limiting process of the N-BSEM.

2.3.3. *As a limit process of the N-BSEM*

In the N-BSEM with (2.4), suppose that there exist some functions $\mu_c(t)$ and $\sigma_c(t)$ such that

$$\sum_{k=0}^{n} \mu_k \mathbf{1}_{(T_k, T_{k+1}]}(t) \to \mu_c(t); \quad \sum_{k=0}^{n} \sigma_k \mathbf{1}_{(T_k, T_{k+1}]}(t) \to \sigma_c(t)$$

for each $t > 0$ as $N \to \infty$ while $\sup_{1 \le k \le N} |T_k - T_{k-1}| \to 0$. As is well known in stochastic analysis, it holds that

$$X_t^N \xrightarrow{\mathbb{P}} X_t^\infty := x + \int_0^t \mu_c(u) \, du + \int_0^t \sigma_c(u) \, dW_u, \quad N \to \infty,$$

uniformly on any compact set for t (ucp-topology). See Protter (2004).

Although there are many studies on the hitting time distribution for such a general diffusion process, only a few models give explicit formulas for the corresponding mortality functions.

2.4. Explicit mortality functions

Suppose X^c satisfies

$$X_t^c = x_c + \int_0^t U(s, \mu_c) \, ds + \int_0^t V(s, \sigma_c) \, dW_s, \tag{2.6}$$

and that there exists a constant $\kappa_c \neq 0$ such that

$$\frac{M(t, \mu_c)}{S(t, \sigma_c)} \equiv \kappa_c, \quad \text{for all } t \ge 0, \tag{2.7}$$

where

$$M(t, \mu_c) = \int_0^t U(s, \mu_c) \, ds; \quad S(t, \sigma_c) = \frac{1}{2} \int_0^t V^2(s, \sigma_c) \, ds.$$

This special SEM gives us an explicit form of the mortality function as follows.

Theorem 2.1 (Molini *et al.*, 2011). *Consider the mortality function of an Id-SEM satisfying the condition (2.7) and put*

$$q_c(t, \theta) := \mathbb{P}(\tau^c \le t), \quad t \ge 0,$$

where $\theta = (x_c, \mu_c, \sigma_c, \kappa_c)$ is the unknown parameter. Then, it holds that

$$1 - q_c(t, \theta) = \int_0^\infty f(z, t|\theta) \, dz, \quad t \ge 0,$$

where

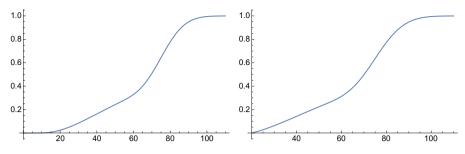$$f(z, t|\theta) = G_\theta(z - x_c, t) - e^{-\kappa_c x_c} G_\theta(z + x_c, t)$$

FIGURE 2: Curves of the unconditional ($S = 0$, left) and conditional ($S = 20$, right) mortality functions of an inhomogeneous diffusion process with (2.9).

and

$$G_\theta(y, t) := \frac{1}{2\sqrt{\pi S(t, \sigma_c)}} \exp\left(-\frac{(y - M(t, \mu_c))^2}{4S(t, \sigma_c)}\right).$$

Consequently, we can construct a parametric family of mortality functions as

$$\mathcal{P}_c = \{q_c(t, \theta) \mid \theta \in \Theta\}, \tag{2.8}$$

where q_c is the function given in Theorem 2.1 and $\Theta \subset \mathbb{R}^m$ with $m := \dim \Theta$.

Figure 2 shows the examples of (conditional) mortality functions

$$q_c(t, \theta \mid S) := \frac{\mathbb{P}(S < \tau^c \leq t)}{\mathbb{P}(\tau^c > S)}, \quad t \geq 0,$$

of an inhomogeneous diffusion process (2.6) with coefficients

$$U(t, \mu_c) = \alpha_c + \beta_c(t - T)^{\gamma_c} \mathbf{1}_{\{t > T\}}, \tag{2.9}$$

with parameters $\mu_c = (\alpha_c, \beta_c, \gamma_c) = (-7.6, -0.009, 2.7)$ and $\kappa = -0.002$. The left curve is the unconditional version ($S = 0$) and the right one is conditional with $S = 20$, which will be used in the data analysis later. Compare it to, for example, Figure 3, the empirical mortality function of Swedish 1830 birth-cohort.

3. STATISTICAL INFERENCE FOR ID-SEM

3.1. Empirical conditional mortality function

Since we cannot observe the path of X^c in reality, estimating the unknown parameters in the SE process will not be straight forward. However, we can see "in principle" the many observations of the time of death τ^c:

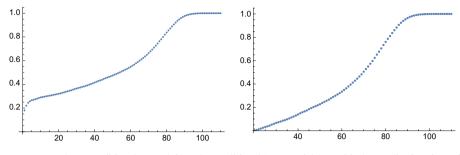$$\tau_1^c, \tau_2^c, \ldots, \tau_{n^c}^c,$$

FIGURE 3: The unconditional $S = 0$, left) and a conditional ($S = 20$, right) empirical mortality functions of the 1830-birth-cohort in Sweden.

where τ_i^c is the time of death of the ith human and n^c is the population in cohort c. Let

$$q_c(t|S) := \mathbb{P}(\tau^c \leq t | \tau^c > S) = \frac{q_c(t)}{1 - q_c(S)}$$

be the true conditional mortality function for a S year old in cohort c. If we observe the actual time of death: τ_i ($i = 1, 2, \ldots, n_c$), we could compute the empirical version of $q_c(\,\cdot\,|S)$ as follows:

$$\widehat{q}_c(t|S) := \frac{\displaystyle\sum_{i=1}^{n_c} \mathbf{1}_{\{S < \tau_i^c \leq t\}}}{\displaystyle\sum_{i=1}^{n_c} \mathbf{1}_{\{\tau_i^c > S\}}}, \tag{3.1}$$

Although it might be hard to obtain such individual data, we can make an empirical version $\widehat{q}_c(t|S)$ using life tables from the Human Mortality Database (HMD) (https://www.mortality.org/) as follows: Let $q_x^{(c)}$ be the mortality rate for 1 year of age x in "calendar year" c, and let q_x^c be the mortality rate for 1 year of age x with "birth year" c (cohort). We note that, in the life table of the calendar year c (from HMD), $q_0^{(c)} = q_0^c$, $q_1^{(c)} = q_1^{c-1}, \ldots, q_\omega^{(c)} = q_\omega^{c-\omega}$, where ω is the final age in the life table. Then, we have that $q_k^c = q_k^{(c+k)}$ for $k = 2, 3, \ldots, \omega$ (e.g., $\omega = 110$ in HMD). From these, we can compute the empirical mortality function

$$\widehat{q}_c(t|S) = 1 - \mathbb{P}(\tau^c > t | \tau^c > S) = 1 - \prod_{k=S}^{t-1} p_k^c, \tag{3.2}$$

for $t = S + 1, \ S + 2, \ \ldots, \ \omega$.

We show in Figure 3 an example of the unconditional/conditional mortality functions of the Swedish 1830-birth-cohort made from HMD (https://www.mortality.org/). We see that the unconditional curve seems to have two points of inflection, which is hard to fit by our parametric family

in (2.8). However, if we take, for example, $S = 20$, we get a conditional curve with one point of inflection, which could be fitted with a mortality function generated by our diffusion models.

3.2. Least squares estimation for hitting time distribution

Consider a parametric model of the mortality function $q_c(t) = \mathbb{P}(\tau^c \leq t)$:

$$\mathcal{P}_c = \{q_c(t, \theta) \mid \theta \in \Theta\},$$

where $\Theta \subset \mathbb{R}^m$ is a bounded open set and suppose that there exists a true value of the parameter $\theta_0 \in \Theta$ such that

$$q_c(t, \theta_0) = q_c(t),$$

for all $t \geq 0$. Moreover, let,

$$q_c(t, \theta \mid S) := \frac{q_c(t, \theta) - q_c(S, \theta)}{1 - q_c(S, \theta)},$$

which is a parametric model for $\mathbb{P}(\tau^c \leq t \mid \tau^c > S)$.

By the Glivenko–Cantelli theorem in the classical theory of statistics, it is easy to see for $\widehat{q}_c(t \mid S)$ with the expression (3.1) that

$$\sup_{t \in [0, \infty)} |\widehat{q}_c(t \mid S) - q_c(t, \theta_0 \mid S)| \to 0 \quad a.s., \quad n_c \to \infty, \tag{3.3}$$

if $q_c(S) < 1$. Therefore, we shall estimate the parameter θ_0 by fitting our parametric model to the empirical version in terms of the least squares error.

Definition 3.1. *For a given S with $0 < q_c(S) < 1$, and $S < t_1 < t_2 < \cdots < t_d \leq \omega$,*

$$\widehat{\theta}_n := \arg \min_{\theta \in \Theta} \sum_{i=1}^{d} |q_c(t_i, \theta \mid S) - \widehat{q}_c(t_i \mid S)|^2,$$

where Θ is an open bounded subset of $\mathbb{R}^m$ and $\overline{\Theta}$ is the closure. The index $n := n_c$ is actually implicit in the estimator since (3.1) is computed the same way as in (3.2) in practice.

Theorem 3.1. *Suppose $q_c(t, \cdot \mid S)$ belongs to $C^1(\Theta)$ for each t and S. Moreover, for a given $S > 0$, suppose the following identifiability condition holds true:*

$$q_c(t_i, \theta \mid S) = q_c(t_i, \theta_0 \mid S) \text{ for } i = 1, 2, \ldots, d \quad \Rightarrow \quad \theta = \theta_0. \tag{3.4}$$

Then, we have a weak consistency of the least square error $\widehat{\theta}_n$:

$$\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \quad n \to \infty.$$

Proof. Note that $\widehat{\theta}_n$ is an M-estimator for the contrast function

$$\ell_n(\theta) := \sum_{i=1}^{d} |q_c(t_i, \theta|S) - \widehat{q}_c(t_i|S)|^2 ,$$

and that, by (3.3), it holds for each $\theta \in \Theta$ that

$$\ell_n(\theta) \to \ell(\theta) := \sum_{i=1}^{d} |q_c(t_i, \theta|S) - q_c(t_i, \theta_0|S)|^2 \quad a.s., \quad n \to \infty.$$

According to the M-estimation theory (e.g., van der Vaart 1998, Theorem 5.7), the proof of consistency is complete if we show the following (a) and (b). There is a function $\ell : \Theta \to \mathbb{R}$ such that

(a) $\sup_{\theta \in \Theta} |\ell_n(\theta) - \ell(\theta)| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty$;

(b) For any $\epsilon > 0$, $\inf_{\theta:|\theta - \theta_0| > \epsilon} |\ell(\theta)| > \ell(\theta_0)$.

Note that (b) is clear with $\ell(\theta_0) = 0$ by the identifiability condition, so we will now show (a).

We consider ℓ_n as a random map $\ell_n : \Omega \to (C(\overline{\Theta}), \|\cdot\|)$, where $\|\cdot\|$ is the sup norm over $\overline{\Theta}$, by extending the domain of ℓ_n to $\overline{\Theta}$ by its continuity on Θ. Then, the uniform tightness of the $C(\overline{\Theta})$-valued random element ℓ_n implies a weak convergence $\ell_n \xrightarrow{\mathcal{D}} \ell$ in $C(\overline{\Theta})$, which is equivalent to (a). A well-known tightness criterion is given by

$$\sup_n \mathbb{E} \left[\sup_{\theta \in \overline{\Theta}} |\partial_\theta \ell_n(\theta)| \right] < \infty,$$

and this is clear since q_c and $\widehat{q}_c$ are bounded by 1 and $\partial_\theta q_c$ is bounded on $\overline{\Theta}$ by the assumption. Hence, the proof is complete. $\qquad \square$

Theorem 3.2. *Suppose the same assumptions as in Theorem 3.1. Moreover, suppose $q_c \in C^2(\Theta)$ and Θ is a convex subset of $\mathbb{R}^m$. Then,*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} R_d^{-1} Q_d \cdot N_d(0, \Sigma), \quad n_c \to \infty$$

where

$$Q_d = (\partial_\theta q_c(t_1, \theta_0|S), \ldots, \partial_\theta q_c(t_d, \theta_0|S)) \in \mathbb{R}^m \otimes \mathbb{R}^d,$$

$$R_d = \sum_{i=1}^{d} \left[(\partial_\theta^\top q_c)(\partial_\theta q_c)(t_i, \theta_0|S) + q_c(t_i, \theta)\partial_\theta^2 q_c(t_i, \theta_0|S) \right] \in \mathbb{R}^m \otimes \mathbb{R}^m,$$

and the variance-covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq m}$ is given by

$$\sigma_{ij} = \frac{1}{\overline{q}_c^2(S)} \Lambda(t_i, t_j) + \frac{\overline{q}_n(t_i)}{\overline{q}_c^4(S)} \Lambda(S, S) + \frac{\overline{q}_c(t_i)}{\overline{q}_c^3(S)} \Lambda(t_i, S) + \frac{\overline{q}_c(t_j)}{\overline{q}_c^3(S)} \Lambda(t_j, S)$$

with $\Lambda(x, y) = q_c(x \wedge y) - q_c(x) q_c(y)$ and $\overline{q}_c = 1 - q_c$.

Proof. We use the notation $\partial_\theta = (\partial_{\theta_1}, \dots, \partial_{\theta_m})^\top$ and $\partial_\theta^2 = \partial_\theta \partial_\theta^\top$.

For the proof, we shall use the same notation as the previous proof of Theorem 3.1. Moreover, for simplicity, we suppose the M-estimator $\widehat{\theta}_n$ belongs to the interior of the parameter space Θ as usual. Then, we can assume, by the usual argument, that $\partial_\theta \ell_n(\widehat{\theta}_n) = 0$, and that $\det \partial_\theta^2 \ell_n(\theta) \neq 0$ around θ_0 without loss of generality because the probability of those becomes zero when $n_c \to \infty$.

Under the above assumptions, we see from the mean value theorem that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \left[\partial_\theta^2 \ell_n(\theta_n^*)\right]^{-1} \cdot \sqrt{n} \partial_\theta \ell_n(\theta_0),$$

where we note that

$$\sqrt{n} \partial_\theta \ell_n(\theta_0) = 2 \sum_{i=1}^d \sqrt{n} \left(\widehat{q}_c(t_i|S) - q_c(t_i, \theta_0|S)\right) \partial_\theta q_c(t_i, \theta_0|S),$$

$$\partial_\theta^2 \ell_n(\theta) = 2 \sum_{i=1}^d \left[(\partial_\theta^\top q)(\partial_\theta q)(t_i, \theta|S) + q(t_i, \theta) \partial_\theta^2 q(t_i, \theta|S)\right].$$

Note that $\widehat{q}_c(t_i|S)$ is the empirical distribution that is consistent with $q_c(t_i|S)$:

$$\widehat{q}_c(t_i|S) = \frac{\widehat{F}_n(t_i) - \widehat{F}_n(S)}{1 - \widehat{F}_n(S)}, \quad q_c(t_i|S) = \frac{F(t_i) - F(S)}{1 - F(S)},$$

where F is the distribution function of τ_i^c and

$$\widehat{q}_n(t) := \frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{1}_{\{\tau_i^c \leq t\}}$$

and that it follows from the Donsker-type theorem that, for any $x_i \in \mathbb{R}$,

$$\left(\sqrt{n}(\widehat{q}_n(x_1) - q_c(x_1)), \dots, \sqrt{n}(\widehat{q}_n(x_K) - q_c(x_K))\right) \xrightarrow{\mathcal{D}} (Z_{x_1}, \dots, Z_{x_k}) \sim N_K(0, \Lambda),$$

as $n_c \to \infty$, where $\Lambda = (\Lambda(x_i, x_j))_{1 \leq i,j \leq K}$ with $\Lambda(x_i, x_j) = q_c(x_i \wedge x_j) - q_c(x_i) q_c(x_j)$ (see van der Vaart 1998, Theorem 19.3). Hence, it follows that

$$\sqrt{n}(\widehat{q}_c(t_i|S) - q_c(t_i|S)) = \frac{\sqrt{n}(\widehat{q}_n(t_i) - q_c(t_i)) - \sqrt{n}(\widehat{q}_n(S) - q_c(S))}{1 - \widehat{q}_n(S)}$$

$$+ (q_c(t_i) - q_c(S))\sqrt{n} \left(\frac{1}{1 - \widehat{q}_n(S)} - \frac{1}{1 - q_c(S)}\right)$$

$$\xrightarrow{\mathcal{D}} \frac{1}{\overline{q}_c(S)} Z_{t_i} + \frac{\overline{q}_c(t_i)}{\overline{q}_c^2(S)} Z_S$$

jointly for $i = 1, 2, \ldots, d$, and therefore,

$$\sqrt{n}\partial_\theta \ell_n(\theta_0) \xrightarrow{\mathcal{D}} 2 \sum_{i=1}^{d} \left[\frac{1}{\overline{q}_n(S)} Z_{t_i} + \frac{\overline{q}_n(t_i)}{\overline{q}_n^2(S)} Z_S \right] \partial_\theta q_c(t_i, \theta_0 | S)$$

$$= 2Q_d \cdot N_d(0, \Sigma),$$

where $\Sigma = (\sigma_{ij})_{1 \le i,j \le m}$ is given by

$$\sigma_{ij} = \frac{1}{\overline{q}_n^2(S)} \Lambda(t_i, t_j) + \frac{\overline{q}_n(t_i)}{\overline{q}_n^4(S)} \Lambda(S, S) + \frac{\overline{q}_c(t_i)}{\overline{q}_c^3(S)} \Lambda(t_i, S) + \frac{\overline{q}_c(t_j)}{\overline{q}_c^3(S)} \Lambda(t_j, S)$$

Moreover, since $\theta_n^* \xrightarrow{\mathbb{P}} \theta_0$ by the consistency of $\widehat{\theta}_n$, we see from the continuous mapping theorem that

$$\partial_\theta^2 \ell_n(\theta_n^*) \xrightarrow{\mathbb{P}} \sum_{i=1}^{d} \left[(\partial_\theta^\top q_c)(\partial_\theta q_c)(t_i, \theta_0 | S) + q_c(t_i, \theta_0) \partial_\theta^2 q_c(t_i, \theta_0 | S) \right] = R_d.$$

This completes the proof.   □

Remark 3.1. *Our structural approach meets the consistency and asymptotic normality for a suitable parametric SEM. This is a significant advantage over the Lee–Carter type reduced form model of mortality rate, where the consistency of the estimator is difficult to establish (see Leng and Peng 2016 for details).*

4. DATA ANALYSIS WITH ID-SEM

4.1. Id-SEM with the exact mortality function

Here, we try the inhomogeneous diffusion process (2.6) with the (2.7) condition for the Id-SEM process to specify the mortality functions. We use life tables of several countries from HMD (https://www.mortality.org/) and estimate the model parameters as described in Section 3.

We choose the three models A–C as candidates of coefficient $U(t, \mu_c)$ in the diffusion model with parameter $\mu_c = (\alpha_c, \beta_c, \gamma_c)$:

Model A: $U(t, \mu_c) = \alpha_c$ (toy: drifted Brownian motion);
Model B: $U(t, \mu_c) = \alpha_c + \beta_c(t - T)^{\gamma_c} \mathbf{1}_{\{t > T\}}$ (polynomial);
Model C: $U(t, \mu_c) = \alpha_c + \beta_c \exp(\gamma_c(t - T)) \mathbf{1}_{\{t > T\}}$ (exponential);

and the diffusion coefficient is given by $V^2(t) = \frac{2}{\kappa_c} U(t, \mu_c)$. We try the least squares method described in the previous section with $S = 20$ in each model. Here, Model A is the toy model described in Section 2.1. We can expect that the mortality function of Model A will not have a rapid increase in old age after, say, 50 years. Model B can improve the disadvantage of Model A with a

polynomial growth function after change point T. Model C is an exponential growth model.

Throughout the analyses in this section, we fix the initial value $x_c = 1000$ to satisfy the identifiability condition (3.4) in Theorem 3.1. Although we can take x_c arbitrarily, we chose $x_c = 1000$ to make the order of estimated values of other parameters be numerically appropriate. We note that, for any x_c, there is some μ_c that gives the same fitting in terms of the least squared error in our model. That is, the initial value x_c is not that important in this SEM. Thus, the parameter $\theta = (\alpha_c, \beta_c \gamma_c, \kappa_c)$ is to be estimated and predicted for the future model.

4.1.1. *Swedish mortality data*

Swedish mortality data are good for testing the prediction for the distant future because the HMD (https://www.mortality.org/) has the long-term life tables of 20–110 years from 1751 to 2016. We apply our Id-SEM to the data from 1801 to 1830 birth-cohort and try to predict the mortality functions of the cohort after 10, 30, 50, and 70 years.

First, we shall try to estimate the parameters from the historical mortality functions with the change point $T = 50$. That is, assuming that we are now in 1940, we estimate the parameters of

- 1801-birth-cohort by life tables of 1821–1911;
- 1802-birth-cohort by life tables of 1822–1912;
  ⋮
- 1830-birth-cohort by life tables of 1850–1940,

Figure 4 shows an example of the estimated curves and the least squared error in models A–C for the $c =$ 1830 birth-cohort. As expected, Model A does not follow the empirical curves. Model B performs better than Model A but Model C is the best in terms of the mean squared error of all cohorts (1801–1830). Hence, we use the exponential growth model (Model C) to predict the future mortality functions.

The results of all the parameter estimates in Model C are given in Figure 5. From the plots of the estimated values, we use linear regression for predicting the value of γ, and the exponential functions for α_c, β_c, and κ_c because they seem to be increasing but their sign should be negative. We predict the future parameters using these regression functions and show the prediction curves after 10, 30, 50, and 70 years in Figure 6 with the prediction (LSE) errors in Table 1.

From the results, we observe that our Id-SEM can predict future mortality very well even after 50 years though there is a small gap after 70 years.

4.1.2. *Dutch mortality data*

Next, we apply our Models A–C to Dutch data of 1841–1870 birth-cohorts. In this example, we chose $S = 20$ and change point $T = 60$, which is selected so that the least square error fitting of the models is as good as possible.

TABLE 1

SWEDEN: THE MEAN SQUARED PREDICTION ERRORS BETWEEN THE EMPIRICAL
AND THE PREDICTED MORTALITY FUNCTIONS.

Cohort	Model A	Model B	Model C
1840 (10 years)	0.457803	1.26366	0.0091356
1860 (30 years)	0.670725	3.3604	0.0382479
1880 (50 years)	1.0998	5.28708	0.0180587
1900 (70 years)	1.1654	6.20919	0.1786

Model	MSE
A (toy)	0.408821
B (pol.)	0.003074
C (exp.)	**0.000317**

FIGURE 4: Swedish mortality function of 1830-birth-cohort: estimated curve (solid), empirical curve (dashed), and the least squared error of each estimated curve.

According to Figure 7 with the table of mean squared errors, we see that Model C is the best in 1870-cohort; it was the same as in the other cohorts of 1841–1869. From the results of estimated parameters in 1841–1870 birth-cohorts, see Figure 8, we predict the future parameters using nonlinear regression as in the previous section and show the mortality functions of 10, 20, and 30 years after 1870 with the obtained (empirical) mortality data in Figure 9 and the prediction errors in Table 2. We also see in this example that our model can predict future mortality very well.
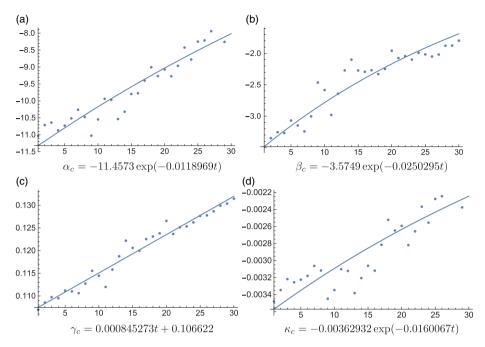
FIGURE 5: Parameter estimates for (a) α_c, (b) β_c, (c) γ_c, and (d) κ_c in Model C from Swedish mortality data for 30 years ($c = 1801$–1830) with results of (non) linear regressions. The horizontal axis is t, for which $c = 1800 + t$.
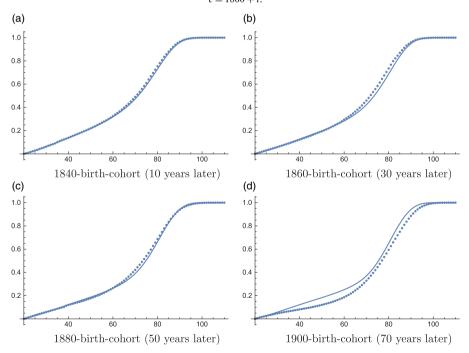

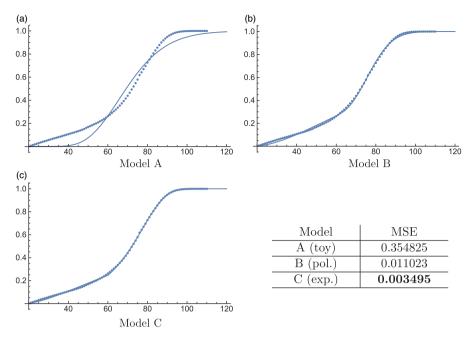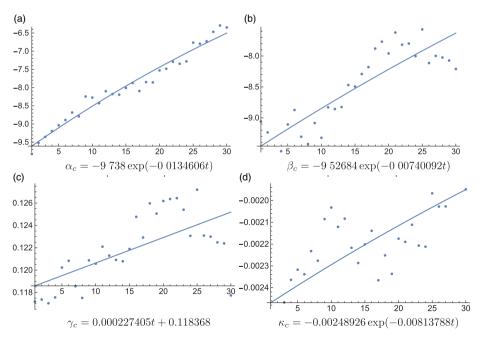FIGURE 6: Predicted Swedish mortality functions and the obtained mortality curves.

FIGURE 7: Dutch mortality function of 1870-birth-cohort: the estimated curve (solid) and empirical (dashed), and the mean squared error (MSE) of each estimated curve.

Model	MSE
A (toy)	0.354825
B (pol.)	0.011023
C (exp.)	**0.003495**

FIGURE 8: Parameter estimates for (a) α_c, (b) β_c, (c) γ_c, and (d) κ_c in Model C from Dutch mortality data for 30 years ($c = 1841$–1870) with results of (non) linear regressions. The horizontal axis is t, for which $c = 1840 + t$.

TABLE 2

THE NETHERLANDS: THE MEAN SQUARED PREDICTION ERRORS BETWEEN
THE EMPIRICAL AND THE PREDICTED MORTALITY FUNCTIONS.

Cohort	Model A	Model B	Model C
1880 (10 years)	0.330259	0.0197578	0.0220831
1890 (20 years)	0.329853	0.0261606	0.0623656
1900 (30 years)	0.256442	0.0287624	0.0995795

FIGURE 9: Predicted Dutch mortality functions and the realized mortality curves.

4.1.3. *Japanese mortality data*

We next apply our models to Japanese mortality data.

As Japan currently has a rapidly aging population, it is very important to predict the future mortality function, especially in old age. Since there is not much Japanese mortality data in the HMD (https://www.mortality.org/), we cannot confirm the accuracy of our predictions for the distant future. However, from the good prediction results for the Swedish and Dutch data, we expect that our model would yield good predictions for the Japanese data too.

We use the Japanese mortality data from 1947 to 2017. First, we estimate the parameters in 1897–1907 birth-cohort (for 11 cohorts) with $S = 50$ and $T = 70$ up to 90 year old, and predict the mortality of age 50 years in the cohort of 10–40 years after 1907.
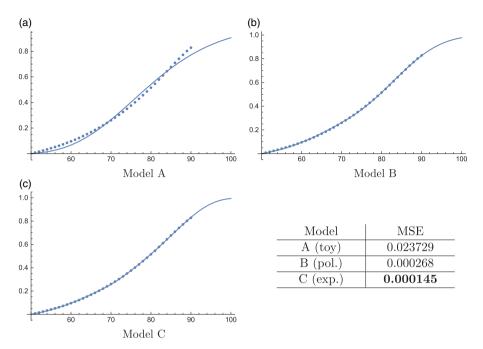
FIGURE 10: Japanese mortality function of 1907-birth-cohort: the estimated curve (solid), empirical curve (dashed), and the mean squared error (MSE) of each estimated curve.
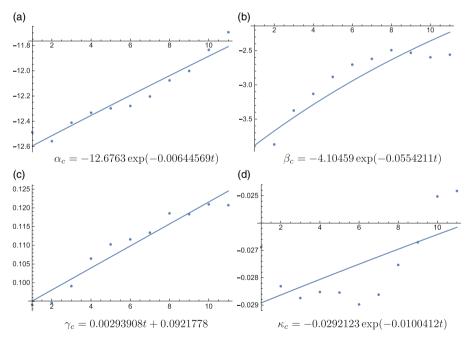
Model	MSE
A (toy)	0.023729
B (pol.)	0.000268
C (exp.)	**0.000145**

$$\alpha_c = -12.6763 \exp(-0.00644569t)$$

$$\beta_c = -4.10459 \exp(-0.0554211t)$$

$$\gamma_c = 0.00293908t + 0.0921778$$

$$\kappa_c = -0.0292123 \exp(-0.0100412t)$$

FIGURE 11: Parameter estimates for (a) α_c, (b) β_c, (c) γ_c, and (d) κ_c in Model C from Japanese mortality data for 11 years ($c = 1897$–1907) with the results of (non) linear regressions. The horizontal axis is t, for which $c = 1896 + t$.

JAPAN: THE MEAN SQUARED PREDICTION ERRORS BETWEEN THE EMPIRICAL
AND THE PREDICTED MORTALITY FUNCTIONS.

Cohort	Model A	Model B	Model C
1917 (10 years)	0.0289017	0.00893398	0.00682046
1927 (20 years)	0.0186373	0.0279498	0.00422352
1937 (30 years)	0.00564805	0.00286146	0.00194083
1947 (40 years)	0.00849966	0.00242087	0.000961335

FIGURE 12: Predicted Japanese mortality functions and the obtained mortality curves.

Figure 10 shows the least square error fitting of SEMs A–C for old age in the 1907 birth-cohort. We also observe that Model C always gives a better fit to the empirical data for 1897–1906. The results of parameter estimation and nonlinear regression in each cohort are given in Figure 11. Based on these, the predicted mortality functions are plotted in Figure 12 and the prediction errors in Table 3. The predicted curve in each cohort seems to fit the obtained mortality although the actual (data) are up to 2017. We observe that the death rate at each age is decreasing in each cohort, which indicates that Japanese society is going to age more and more.

4.1.4. *English and Welsh mortality data*
Lastly, we use data from England and Wales. In this respect, we find a different feature for the transition of the parameters. We tried a preliminary analysis
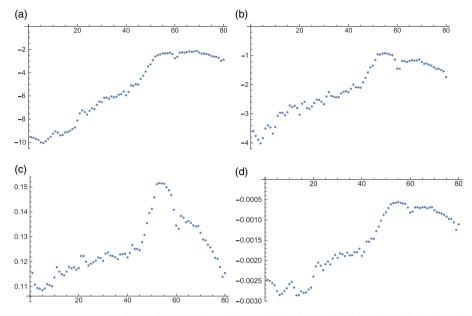
FIGURE 13: Parameter estimates for (a) α_c, (b) β_c, (c) γ_c, and (d) κ_c in Model C from the English and Welsh mortality data for 80 years ($c = 1841$–1920). The horizontal axis is t, for which $c = 1840 + t$. There is a "structural change" around $c = 1890$ ($t = 50$).

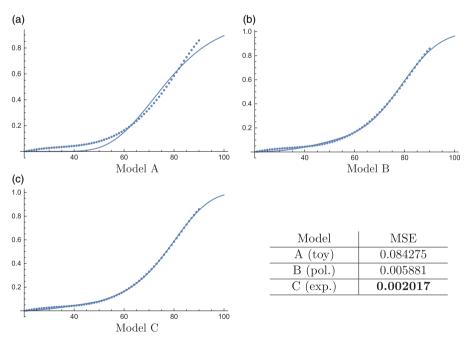| Model	MSE
A (toy)	0.084275
B (pol.)	0.005881
C (exp.)	**0.002017**

FIGURE 14: English and Welsh mortality function of 1920-birth-cohort: the estimated curve (solid), empirical curve(dashed), and the mean squared error (MSE) of each estimated curve.
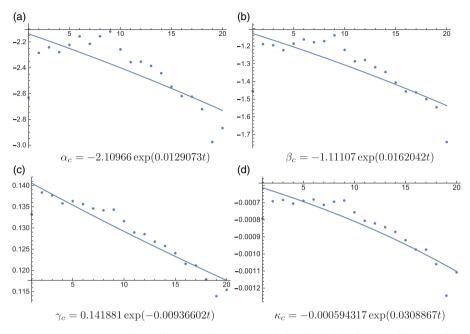
FIGURE 15: Parameter estimates for (a) α_c, (b) β_c, (c) γ_c, and (d) κ_c in Model C from English and Welsh mortality data for 20 years ($c = 1901$–1920) with the results of (non) linear regressions. The horizontal axis is t, for which $c = 1900 + t$.

with Model C for the data of the birth-cohorts 1841–1920; see Figure 13. We observed a "structural change" around 1890 though we do not know the reason for it.

Therefore, we will use the data from 1901 to 1920 birth-cohort (for 20 cohorts) in the least square error fitting of the mortality function with $S = 20$ and $T = 50$ up to 90 year old, and predict the mortality of age 20 years in the cohort of 10–30 years after 1920.

The results of fitting Models A–C for the 1920 birth-cohort are given in Figure 14. We also observe that Model C again gave a better fit to the data for 1901–1920, and the results of its parameter estimation and nonlinear regression for each cohort are given in Figure 15.

Remark 4.1. *In order to capture the "structural change" around 1890 birth-cohort, we had also tried to predict parameter changes by time series ARIMA model (with the model selection by AIC) including whole the term 1841–1920. However, the predicting results were very poor. Hence we will only show the results by nonlinear regression in this paper. How to predict such a structural change is an important issue in the future.*

We predict the mortality function of 20 years old of the future cohort $c = 1930$ (10 years later), 1940 (20 years later), and 1950 (30 years later). The mean squared errors for the prediction are given in Table 4. We find that Model

TABLE 4

ENGLAND AND WALES: THE MEAN SQUARED PREDICTION ERRORS BETWEEN
THE EMPIRICAL AND THE PREDICTED MORTALITY FUNCTIONS.

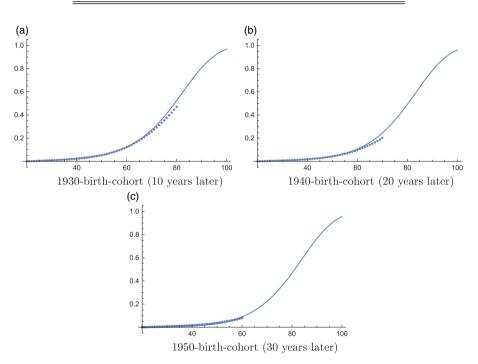| Cohort	Model A	Model B	Model C
1930 (10 years)	0.0624429	0.0146555	0.0149824
1940 (20 years)	0.0251397	0.00568472	0.00735449
1950 (30 years)	0.0106258	0.00108093	0.000829724

FIGURE 16: Predicted English and Welsh mortality functions by Model C and the obtained mortality curves.

C would be again a better choice for the prediction, and the predicted mortality functions are given in Figure 16.

Although the predicted curve is slightly bigger than the actual around age 70 years, it seems to fit the data in each cohort very well. We again observe as in the previous results for Japan that the death rate at each age is decreasing in the cohort, which indicates that the English society is also going to age more in future.

4.2. Comparison with the Lee–Carter model

Finally, we close this section by comparing our model to the classical Lee–Carter model given in (1.1). Although this comparison may be unfair since the Lee–Carter model (1.1) does not include cohort effect, we should attempt it because the Lee–Carter model is the most standard model in practice.
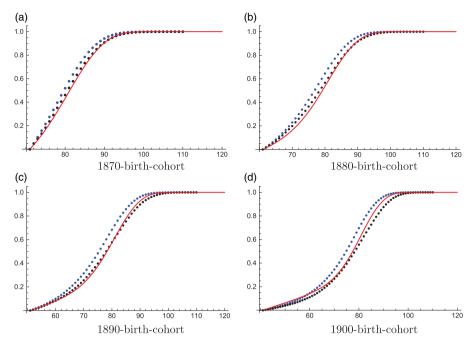
FIGURE 17: Predicted mortality functions: Lee–Carter (blue dots) versus Id-SEM (red solid line). The black dots are the empirical (actual) mortality function.

We will use the Swedish data of 1821–1940 life tables from HMD (https://www.mortality.org/) as well as Section 4.1.1. First, we use SEM (Model C) and estimate the parameters of 1801–1830 birth-cohorts (20–110 years old), and predict the mortality for age: (a) 70–110 years in the 1870-birth-cohort; (b) 60–110 years in the 1880 birth-cohort; (c) 50–110 years in the 1890 birth-cohort; and (d) 40–110 years in the 1900 birth-cohort. Assuming that we are now in 1940, we estimate the parameters of

- 1801-birth-cohort by life tables of 1821–1911;
- 1802-birth-cohort by life tables of 1822–1912;
  ⋮
- 1830-birth-cohort by life tables of 1850–1940,

For example, a human who was born in 1870 would be 70 years old in 1940, so (a) is the prediction of mortality of 1870 birth-cohort in 1941–1980, among others.

For the prediction by the Lee–Carter model (1.1), we use the same procedure as in Lee and Carter (1992) to estimate and predict the parameters, and the life tables of 1901–1940, which are enough for prediction by this model since the Lee–Carter model does not consider cohort effect.

The results are given in Figure 17. The black dots are the empirical mortality function (actual), and the red solid lines are the predicted mortality functions by our Id-SEM. The blue dots are the results by the Lee–Carter

model. We find that the SEM gives better predictions than Lee–Carter in all the cohorts, especially for older ages.

5. CONCLUDING REMARKS: SUMMARY AND FUTURE RESEARCH

5.1. Summary

In this study, we propose a new approach to cohort-wise mortality prediction. We assume the SE process for people in each cohort and define the time of death as the first hitting time (τ^c) of the SE process to zero level. We call the distribution function of τ^c the *mortality function*, which is useful for the computations of many actuarial quantities and longevity analysis.

This study explains a human's death in terms of *survival energy* but that does not mean that we claim the existence of such an energy. We claim that some hitting time distributions of diffusion processes can give a good parametric family to fit the empirical mortality functions and their predictions. Indeed, we propose a special inhomogeneous diffusion (Id) process as in Section 2.4, which can give an explicit mortality function and show in Section 4.1 that it gives surprisingly good predictions for the distant future if we choose the drift and diffusion coefficients appropriately.

Although we have tried, in some sense, *ad hoc* models for fitting and predictions, our structural approach is worth using and studying further for mortality prediction.

5.2. Future studies

5.2.1. *Model selection and predictions*
There are many things to be considered from the perspective of modeling.

We fix the change point T of the drift coefficient in Id-SEM for all the cohorts in a country. However, this may depend on the cohort in practice and we need to investigate the estimation and prediction of T from the data. Conversely, we should take S in (3.1) to be high to some extent since the mortality function in the younger population, say, 0–5 years old is not likely to be the hitting time distribution function of a diffusion process; you can compare Figures 2 and 3, ($S = 0$). Therefore, the prediction for younger ages is an important problem.

How to choose T (as well as S) is a problem of model selection since it affects the coefficients of the SE processes. Moreover, we should compare several competing models with different coefficients. A standard procedure for model selection is based on *information criteria*. However, we did not discuss the information criteria in this study because of several theoretical problems.

We estimate the parameter in the SEM using the least squares method, for which the Akaike information criteria (AIC) and Bayesian information criteria (BIC) are not available since they have to be based on the maximum likelihood

estimators. For example, in a theoretical (statistical) perspective, AIC cannot be applied even to the Lee–Carter model because the consistency of estimators of parameters in the Lee–Carter model is not ensured; see Remark 3.1. In our case, we need to construct the generalized information criteria (see Konishi and Kitagawa 1996), which would be not easy since our data for τ_i are restricted. In HMD (https://www.mortality.org/), we cannot obtain the actual of each τ_i but we can see the (approximate) empirical mortality functions that are manufactured from the data of death probabilities. This could be improved if the mortality database is arranged in more detail. Otherwise, we still have many technical (mathematical) issues to be solved under our limited sampling situation. This would be an important issue for future studies.

In the prediction of parameter changes, we used nonlinear regressions for simplicity. However, that might be inadequate because the estimated parameters are sometimes unstable and the trends do not seem to be selected nonlinear functions. For the prediction of parameters, it might be possible to use a machine learning method with Gaussian processes, among others. If we can propose a kind of confidence interval for the future parameters, then we can give a confidence interval for the predicted mortality function. Currently, we can only give a confidence interval for estimated mortality functions by the asymptotic normality results for the estimators: Theorem 3.2.

5.2.2. *Some directions of model extension*

Future studies will have to extend our SEM in several directions: one is to extend the coefficient of the SDE (2.3) to be more general, for example, state-dependent model or random coefficient and the other is to introduce "jumps" in the SE process such as jump-diffusions or Lévy processes since SE could jump if there is a disease or an accident in their lives.

In those cases, we face a problem such that the hitting time distribution q_c is usually not easy to use. In some special cases, we could use some implicit results, for example, by Hao *et al.* (2013) or Abbring (2012), where an integral-type equation or the Laplace transformation of the hitting time distribution is available. However, in most cases, the mortality function is not explicit. Then, we will face problems in statistical inference, and it could sometimes be computer intensive in the end.

One practicable and strong candidate would be the *Inverse Gaussian process*. For example, Ye and Chen (2014) investigate the process in a context of modeling the degradation of products in random environments. Since "degradation" is an accumulation of damage over time that ultimately leads to a system failure when it crosses a certain threshold, the motivation is very similar to our SE hypothesis. In the inverse Gaussian model, the hitting time distribution can be written in explicit form, so it can be a strong candidate for the SEM.

Although Ye and Chen (2014) also investigate the procedures of parameter estimation, there is an important difference between their model and ours.

That is, we cannot observe the SE process directly while they can observe the degradation process. This makes our problem harder on many statistical issues and model selection. We have proposed one possible strategy in this paper, but further study is required on those points as pointed out in Section 5.2.1.

ACKNOWLEDGMENTS

REFERENCES

ABBRING, J.H. (2012) Mixed hitting-time models. *Econometrica*, **80**(2), 783–819.

BAUER, D., BENTH, F.E. and KIESEL, R. (2012) Modeling the forward surface of mortality. *SIAM Journal on Financial Mathematics*, **3**, 639–666.

BAUKAI, B. (1990) An explicit expression for the distribution of the supremum of brownian motion with a change point. *Communications in Statistics - Theory and Methods*, **19**(1), 31–40.

BIFFIS, E. (2005) Affine processes for dynamic mortality and actuarial valuation. *Insurance: Mathematics and Economics*, **37**, 443–468.

BIFFIS, E., DENUIT, M. and DEVOLDER, P. (2010) Stochastic mortality under measure changes. *Scandinavian Actuarial Journal*, **2010**(4), 284–311.

CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2006) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, **73**, 687–718.

CAIRNS, A.J.G., BLAKE, D. and DOWD, K. (2008) Modelling and management of mortality risk: A review. *Scandinavian Actuarial Journal*, **2008**(2–3), 79–113.

CAIRNS, A.J.G., BLAKE, D., DOWD, K., COUGHLAN, G.D., EPSTEIN, D., ONG, A. and BALEVICH, I. (2009) A quantitative comparison of stochastic mortality models using data from England & Wales and the United States. *North American Actuarial Journal*, **13**(1), 1–35.

CHEN, H. and COX, S.H. (2009) Modeling mortality with jumps: applications to mortality securitization. *The Journal of Risk and Insurance*, **76**(3), 727–751.

DAHL, M. (2004) Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics*, **35**, 113–136.

LEE, R.D. and CARTER, L. (1992) Modelling and forecasting the time series of US mortality. *Journal of the American Statistical Association*, **87**, 659–671.

LENG, X. and PENG, L. (2016) Inference pitfalls in Lee-Carter model for forecasting mortality. *Insurance: Mathematics and Economics*, **70**, 58–65.

HAO, X., LI, X. and SHIMIZU, Y. (2013) Finite-time survival probability and credit default swaps pricing under geometric Lévy markets. *Insurance: Mathematics and Economics*, **53**, 14–23.

HUMAN MORTALITY DATABASE: https://www.mortality.org/. Netherlands:2018.10, Sweden:2019.1, Japan:2019.8, England&Wales(civilian):2020.5.

ITO, R. and SHIMIZU, Y. (2019) Cohort-wise mortality prediction under survival energy hypothesis (in Japanese). *Journakl of the Japanese Association of Risk, Insurance and Pensions (JARIP)*, **6**, 17–30.

JARROW, R.A. and TURNBULL, S. (1995) Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, **50**(1), 53–85.

KONISHI, S. and KITAGAWA, G. (1996) Generalized information criteria in model selection. *Biometrika*, **83**, 875–890.

LUNDBERG, F. (1903) *Approximerad Framställning av Sannolikehetsfunktionen, Aterförsäkering av Kollektivrisker*. Uppsala: Almqvist & Wiksell.

MERTON, R.C. (1974) On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, **29**(2), 449–470.

MERTON, R. (1976) Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **3**, 125–144.

MOLINI, A., TALKNER, P., KATUL, G.G. and PORPORATO, A. (2011). First passage time statistics of Brownian motion with purely time dependent drift and diffusion. *Physica A*, **390**, 1841–1852.

PROTTER, P.E. (2004) *Stochastic Integration and Differential Equations*, 2nd edn. Berlin: Springer-Verlag.

RENSHAW, A.E. and HABERMAN, S. (2003) Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, **33**, 255–272.

RENSHAW, A. and HABERMAN, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **38**, 556–570.

SCHOUTENS, W. and CARIBONI, J. (2009) *Lévy Processes in Credit Risk*, Podstow, UK: John Wiley & Sons.

VAN DER VAART, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

YE, Z. and CHEN, N. (2014) The inverse Gaussian process as a degradation model. *Technometrics*, **56**(3), 302–311.

YASUTAKA SHIMIZU (Corresponding author)
Department of Applied Mathematics
Waseda University
Tokyo, Japan
E-Mail: shimizu@waseda.jp

YUKI MINAMI and RYUNOSUKE ITO
Department of Applied Mathematics
Waseda University
Tokyo, Japan
E-Mails: minami-y@moegi.waseda.jp; ryunosuke.i@fuji.waseda.jp