International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Survival energy models for mortality prediction and future prospects

Yasutaka Shimizu[1,*] [ID], Kana Shirai[2], Yuta Kojima[2], Daiki Mitsuda[2] and Mahiro Inoue[2]

[1]Department of Applied Mathematics, Waseda University, Tokyo, Japan and [2]Graduate School of Fundamental Science and Engineering, Waseda University Tokyo, Japan
*Corresponding author. E-mail: shimizu@waseda.jp

Abstract

The survival energy model (SEM) is a recently introduced novel approach to mortality prediction, which offers a cohort-wise distribution function of the time of death as the first hitting time of a "survival energy" diffusion process to zero. In this study, we propose a novel SEM that can serve as a suitable candidate in the family of prediction models. We also proposed a method to improve the prediction in an earlier work. We further examine the practical advantages of SEM over existing mortality models.

1. Introduction

Statistics over the last few decades demonstrate an increase in life expectancy in many countries. For example, in Japan, the life expectancy in 2020 was 85 years, whereas it was 60 years in 1950. Such a rapid change in longevity is called the "*Longevity Revolution*". This trend confers selectivity and value to human life for individuals. However, this gives rise to several medical, economic, and social welfare problems. For instance, the Japanese financial crisis involving the national pension system is a pressing matter. The prediction of mortality is becoming a critical social issue worldwide.

Since the early 20th century, numerous authors have studied mortality prediction, and a methodology has already been established. Most mortality models treat "death" as the first event of a time-inhomogeneous Poisson process. Let T_x be the remaining lifetime of an individual of age x. It is assumed that

$$\mathbb{P}(T_x > t + 1 \mid T_x > t) = \exp\left(-\int_t^{t+1} \mu(x, s)\, \mathrm{d}s\right),$$

where $\mu(x, t)$ is a (possibly stochastic) intensity function called the *force of mortality* in the insurance context. Previous studies have derived models for $\mu(x, t)$. For instance, certain deterministic mortality models, such as the Gompertz, Makeham, and Heligman-Pollard laws, were introduced in earlier years; More recently, numerous stochastic mortality models have been proposed, such as those developed by Olivieri Olovieri (2001), Biffis (2005), Cairns *et al.* (2006b), Hainaut and Devolder Hainaut and Devolder (2008), Biffis *et al.* (2010), Blackburn and Sherris (2013). Moreover, by assuming that $\mu(x, \cdot)$ is constant between $(t, t + 1]$, say $m(x, t)$, allows for modeling of the mortality $m(x, t)$. This approach corresponds to many established classical models, such as the Lee-Carter (1992), Renshaw-Gaberman (2006), and CBD models (Cairns *et al.*, 2008, 2009), among others. We refer to these approaches as *reduced-form approaches*, because they consider death just as a stochastic event.

Shimizu *et al.* (2020) proposed a *structural approach* under the "survival energy hypothesis", which assumes the existence of survival energy for human beings, and death occurs when the energy dissipates. Shimizu *et al.* (2020) used inhomogeneous diffusion (ID) processes as the cohort-wise survival energy model (SEM), such as $X^c = (X_t^c)_{t \geq 0}$ with cohort c, called *ID-SEM*:

$$X_t^c = x_c + \int_0^t U_c(s, \vartheta_c)\, \mathrm{d}s + \int_0^t V_c(s, \vartheta_c)\, \mathrm{d}W_s, \tag{1.1}$$

where x_c is a positive constant corresponding to the initial survival energy, U_c and V_c are deterministic functions on $\mathbb{R}_+ \times \Theta$ with the parameter space given below, and W is the Wiener process:

$$U_c(t, \vartheta_c) = \alpha_c + \beta_c \exp\left(\gamma_c(t - T_c)\right) \mathbf{1}_{\{t \geq T_c\}};$$

$$V_c(t, \vartheta_c) = \sqrt{\frac{2}{\kappa_c} U_c(t, \vartheta_c)},$$

where T_c is a known parameter called *change point*, at which the trend of survival energy changes drastically. Θ of $\vartheta_c = (\alpha_c, \beta_c, \gamma_c, \kappa_c)$ is given by

$$\Theta \subset \{(\alpha, \beta, \gamma, \kappa) \in \mathbb{R}^4 \mid \alpha < 0,\ \beta < 0,\ \gamma > 0,\ \kappa < 0\}.$$

Under this ID-SEM, they defined the time of death as the first hitting time for X^c to reach zero: $\tau^c := \inf\{t > 0 \mid X_t^c < 0\}$, and illustrated that the *mortality function*,

$$q_c(t) := \mathbb{P}(\tau^c \leq t), \tag{1.2}$$

or more practically, the following *conditional mortality function*:

$$q_c(t|S) := \mathbb{P}(\tau^c \leq t | \tau^c > S) \tag{1.3}$$

for a suitably chosen age S can fit their empirical version computed using data from the human mortality database (HMD) (Human Mortality Database); see also Remark 2.5. This indicates that the SEM can propose an excellent parametric family to predict future mortality functions; nevertheless, it is merely a fictitious assumption.

As described in Shimizu *et al.* (2020), the term "structural approach" follows the structural approach in credit risk analysis. This approach is analogous to the structural approach to "default probability", in which a stochastic process describes the asset price. Default time was defined as the first hitting time to a certain level. These two approaches are mathematically identical, but there is a significant difference from a statistical perspective: we can observe the asset process in a credit risk context unlike the "survival energy." However, we can observe many deaths for many individuals' data, although defaults are not directly observed in default risk calculations (because they are predetermined or assumed to occur before default). We estimated the parameters in the SEM family with careful attention to this point.

The main contribution of this paper is the proposal of a novel SEM in Section 2. The mortality function for ID-SEM is sensitive to the change-point parameter T_c and is difficult to predict for a future cohort because it has no clear trend. To address this issue, we propose a new SEM, IG-SEM, which comprises a simple parametric family without such a threshold and is fully flexible enough to fit the training data without a change point. This is helpful because the mortality function can be written explicitly.

Another contribution is that we propose a methodology to improve long-term future predictions in Section 3. The prediction procedure proposed by Shimizu *et al.* (2020) is satisfactory for mid-term future (approximately 10–30 years future) but not for long-term future (e.g., 40 years future cohort). Occasionally, the predictive mortality function does not fit existing data. Therefore, we implement a two-step procedure: the first step is the same as in Shimizu *et al.* (2020), and in the second step, we refit the predicted mortality to the existing younger generation data using 95% prediction intervals for the parameters. We illustrate that this second step can drastically improve the long-term prediction.

Section 5 discusses some advantages of SEM over classical regression-type models in the reduced-form approach. Although this section only presents a theoretical discussion, there are some ongoing

experimental studies. We refer to, for example, Shirai and Shimizu (2022) for discussing the prediction of full life expectancy via SEM.

Finally, Section 6 introduces the SEM project. which explicitly provides cohort/countrywise mortality functions with parameter values on a website.

2. A new SEM: Inverse Gaussian SEM

Let us introduce some notations to provide a new SEM with an explicit mortality function.

Random variable Y follows an *inverse Gaussian distribution*, that is,

$$Y \sim IG(a, b),$$

with mean a and variance a^3/b if the probability density is given by

$$f_Y(y; a, b) = \sqrt{\frac{b}{2\pi y^3}} \exp\left(-\frac{b(y-a)^2}{2a^2 y}\right), \quad y > 0.$$

Definition 2.1 *(IG-SEM; Inverse Gaussian SEM).* A survival energy process $X^c = (X^c_t)_{t \geq 0}$ follows the *IG-SEM* if

$$X^c_t = x_c - Y^c_t, \quad t \geq 0, \tag{2.1}$$

where $x_c > 0$ is the initial energy and $Y^c \sim IG(\Lambda_c, \sigma_c)$ is an *inverse Gaussian process* with mean function Λ_c and parameter $\sigma_c > 0$; that is, $Y^c_0 = 0$, *a.s.*, and Y^c have independent increments. Moreover, for any $t > s > 0$ and an increasing function Λ_c with $\Lambda_c(0) = 0$, it follows that

$$Y^c_t - Y^c_s \sim IG\left(\Lambda_c(t) - \Lambda_c(s), \sigma_c(\Lambda_c(t) - \Lambda_c(s))^2\right).$$

Remark 2.2. If $\Lambda(t) = t$, then Y is an inverse Gaussian Lévy process that is a spectrally positive pure-jump subordinator. Hence, IG-SEM can include a jump in the path of survival energy, although the path of ID-SEM is continuous.

Such a process is used to model the time of system failure in engineering, where failure occurs at τ^c if the accumulated damage Y^c_t exceeds a certain threshold x_c: $\tau^c = \inf\{t > 0 \mid Y^c_t > x_c\}$, which follows the same idea as our survival energy for human death; refer to Ye and Chen (2014). The following theorem provides the mortality function:

Theorem 2.3. *The mortality function for IG-SEM is given by*

$$q^{IG}_c(t, \vartheta_c) = \Phi\left(\sqrt{\frac{\sigma_c}{x_c}}(\Lambda_{\vartheta_c}(t) - x_c)\right) - e^{2\sigma_c \Lambda_{\vartheta_c}(t)} \Phi\left(-\sqrt{\frac{\sigma_c}{x_c}}(\Lambda_{\vartheta_c}(t) + x_c)\right),$$

where $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz,$

Subsequently, we consider the mean function Λ_{ϑ_c} as follows:

$$\Lambda_{\vartheta_c}(t) = e^{a_c t} + b_c t - 1, \quad \vartheta_c = (a_c, b_c, \sigma_c) \in \Theta,$$

where

$$\Theta \subset \{(a, b, \sigma) \in \mathbb{R}^3 \mid a > 0,\ b > 0,\ \sigma > 0\}.$$

The parameters are estimated in a manner similar to those for ID-SEM, as shown in the data analysis in Section 4.

Remark 2.4. As described in Shimizu *et al.* (2020), we can interpret the parameters and coefficients of the SDE. For example, in ID-SEM, the drift term represents the intrinsic survival power of a human and the diffusion term is affected by the social environment. In IG-SEM, λ_c may correspond to the drift term because it is the mean of the accumulating damage process Y^c, and σ_c may be an environmental parameter because it affects damage variance.

Remark 2.5. We estimate the parameters by least-squares fitting of the "conditional" mortality function $q_c(t|S)$ given in (1.3) to the corresponding empirical version, which can be computed based on the data in the HMD (Human Mortality Database), as explained in Shimizu *et al.* (2023). We often recommend choosing a conditioning age S of approximately 20 years. This is because mortality at young ages is highly volatile and unstable, making it difficult to predict with simple models, such as ours. The value of S must be determined empirically by examining the abundance of the data and mortality rates at young ages, which depends on the country.

3. Modification of estimated mortality functions

Suppose we have estimated the values of ϑ_c for some cohorts $c_1 < c_2 < \cdots < c_m$, say, $\widehat{\vartheta}_{c_1}, \ldots, \widehat{\vartheta}_{c_m}$ via LSE, as in Shimizu *et al.* (2020). We assume that future parameter ϑ_c is determined as follows:

$$\vartheta_c = h(c) + \epsilon_c, \quad \epsilon_c \sim N_p(0, \Sigma_c), \tag{3.1}$$

for the deterministic (unknown) mean function h. Assuming that the estimated parameters $\widehat{\vartheta}_{c_1}, \widehat{\vartheta}_{c_2}, \ldots, \widehat{\vartheta}_{c_m}$ are the realisations of ϑ_{c_i} $(i = 1, \ldots, m)$, we estimate h, parameterized case-by-case, as described in Section 4. Once h is estimated, say $\widehat{h}$, we predict the parameter $\vartheta_{c'}$ for a future cohort c' by

$$\widehat{\vartheta}_{c'} = \widehat{h}(c'), \quad c' > c_m, \tag{3.2}$$

and obtain a *predicted mortality function* (PMF) $q_{c'}(\cdot, \widehat{\vartheta}_{c'})$, as in Shimizu *et al.* (2020). However, in this study, we propose further modifications to improve the prediction.

Based on assumption (3.1), we can construct the α-prediction interval for $\vartheta_{c'}$:

$$\widehat{I}_\alpha^{c',m} := \left[\widehat{\vartheta}_{c'} - z_{\alpha/2}\mathrm{diag}(\widehat{\Sigma}_c^{1/2}), \ \widehat{\vartheta}_{c'} + z_{\alpha/2}\mathrm{diag}(\widehat{\Sigma}_c^{1/2}) \right], \tag{3.3}$$

where $\widehat{\Sigma}_c$ is an estimator of Σ_c in (3.1), and z_α is the $(1 - \alpha)$-percentile of $N(0,1)$; that is,

$$\lim_{m \to \infty} \mathbb{P}\left(\vartheta_{c'} \in \widehat{I}_\alpha^{c',m} \right) = \alpha.$$

(Numerical illustrations for these (95%-) prediction intervals are shown in Figures 1 and 2 in the real data analysis). Using prediction interval $\widehat{I}_\alpha^{c',m}$, we readjust the parameters within the α-prediction interval such that the mortality function can fit the existing (younger) data for cohort c' as follows:

Definition 3.1 (*Modified PMF*). When empirical data $\widehat{q}_{c'}(t|S)$ for $t = t_1, \ldots, t_{d'}$ exist, we reselect the predictor such that

$$\widetilde{\vartheta}_{c'} = \arg\min_{\vartheta \in \widehat{I}_\alpha^{c',m}} \sum_{i=1}^{d'} |q_{c'}(t_i, \vartheta|S) - \widehat{q}_{c'}(t_i|S)|^2, \tag{3.4}$$

where $\widehat{I}_\alpha^{c',m}$ is given by (3.3). We used $q_{c'}(\cdot, \widetilde{\vartheta}_{c'})$ as the final predictive mortality function. We refer to it as the *modified predicted mortality function (MPMF)*.

Later, in certain examples, we compare the direct prediction (3.2) with the above modification (3.4).

4. Data analysis: ID-SEM versus IG-SEM

In this section, we compare ID-SEM and IG-SEM using actual data from the HMD (Human Mortality Database) and illustrate that the MPMF with (3.4) can predict future mortality significantly better than the PMF without modification.

Remark 4.1. Determining the change-point parameter T_c in ID-SEM is difficult. In principle, it should be estimated from data, but this is challenging because the estimated mortality function is susceptible
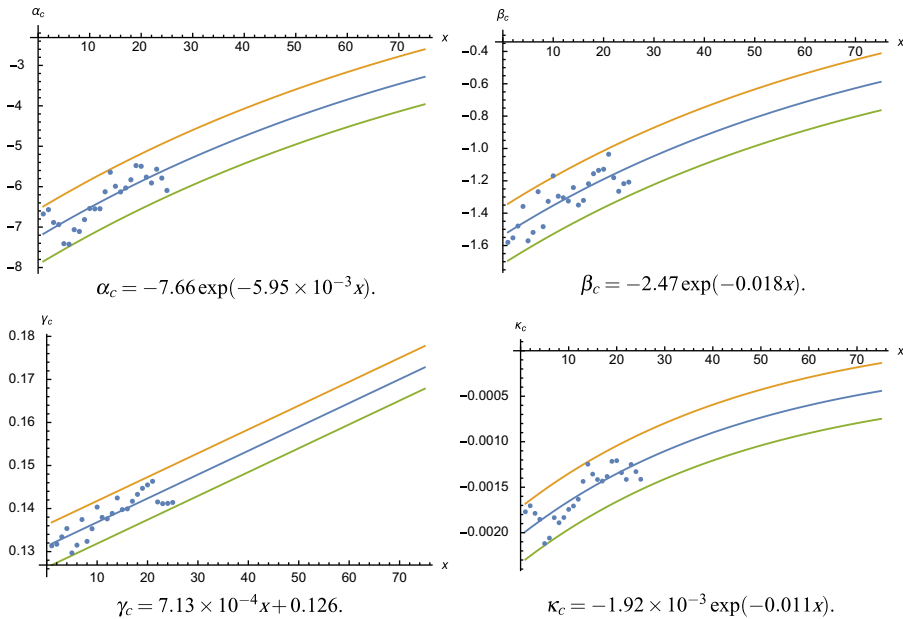
Figure 1. *Estimation of parameters by nonlinear regressions in ID-SEM of Denmark, Female. The x-axis represents $x = c - 1815$; The blue lines are the regression curves. The orange and green curves are upper and lower 95%-prediction bound $(\widehat{I}_{0.95}^{c',25})$, respectively.*

to this parameter. In the following examples, we fix $T_c = 50$, for which ID-SEM can fit the training data relatively well.

4.1 Denmark

The first example is Denmark. We use the following mortality data from the $m = 25$ cohorts:

$$c_1 = 1816 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old;}$$
$$c_2 = 1817 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old;}$$
$$\vdots \qquad\qquad\qquad\qquad \vdots$$
$$c_{25} = 1840 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old,}$$

and suppose that we are in 1951 (because we already have the data of 110 years old of the 1840 birth cohort). Based on this data, we predicted the mortality functions of 20 years old in the future cohorts $c' = 1850$, 1870, and 1890 for females and males, respectively. The predicted age groups for $c' = 1850$, $c' = 1870$, and $c' = 1890$ will be 101 years old, 81 years old, and 61 years old, respectively, based on the assumption that the current year is 1951.

Data analysis was performed using the following procedure.

1. We estimate the parameters in $q_c^{ID}(t, \vartheta_c)$ and $q_c^{IG}(t, \vartheta_c)$ for the data $c = 1816 - 1840$ and obtain the values of the parameters in the future cohorts $c = 1850$, 1870 and 1890 as in Section 3; also refer to Shimizu *et al.* (2020).

   The results for ID-SEM and IG-SEM with the (adjusted) coefficient of determination R^2, $(\overline{R}^2)$ and 95. We will show the tables for R^2 $(\overline{R}^2)$ and the regression curves with the amplitude of the 95%-PI for males, but not the corresponding figures.

Table 1. *The (adjusted) coefficient of determination R^2 ($\overline{R}^2$) for nonliner regression of each parameter with the 95%-prediction intervals (95%-PI). For nonlinear exponential regression, we computed the R^2 ($\overline{R}^2$) by transforming it to the linear regression after taking the logarithm on both sides.*

ID-SEM		
Females	$\alpha_c = -7.66 \exp(-5.95 \times 10^{-3}x)$	$\beta_c = -2.47 \exp(-0.018x)$
(95%-PI)	± 0.663	± 0.347
R^2 ($\overline{R}^2$)	0.477 (0.454)	0.720 (0.708)
Females	$\gamma_c = 7.13 \times 10^{-4}x + 0.126$	$\kappa_c = -1.92 \times 10^{-3} \exp(-0.011x)$
(95%-PI)	$\pm 6.04 \times 10^{-3}$	$\pm 2.38 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.751 (0.740)	0.511 (0.490)
Males	$\alpha_c = -9.26 \exp(-7.25 \times 10^{-3}x)$	$\beta_c = -3.90 \exp(-7.41 \times 10^{-3}x)$
(95%-PI)	± 0.755	± 0.562
R^2 ($\overline{R}^2$)	0.561 (0.542)	0.359 (0.331)
Males	$\gamma_c = 2.98 \times 10^{-4}x + 0.112$	$\kappa_c = -2.70 \times 10^{-3} \exp(-0.0125x)$
(95%-PI)	$\pm 5.03 \times 10^{-3}$	$\pm 4.81 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.432 (0.407)	0.334 (0.305)
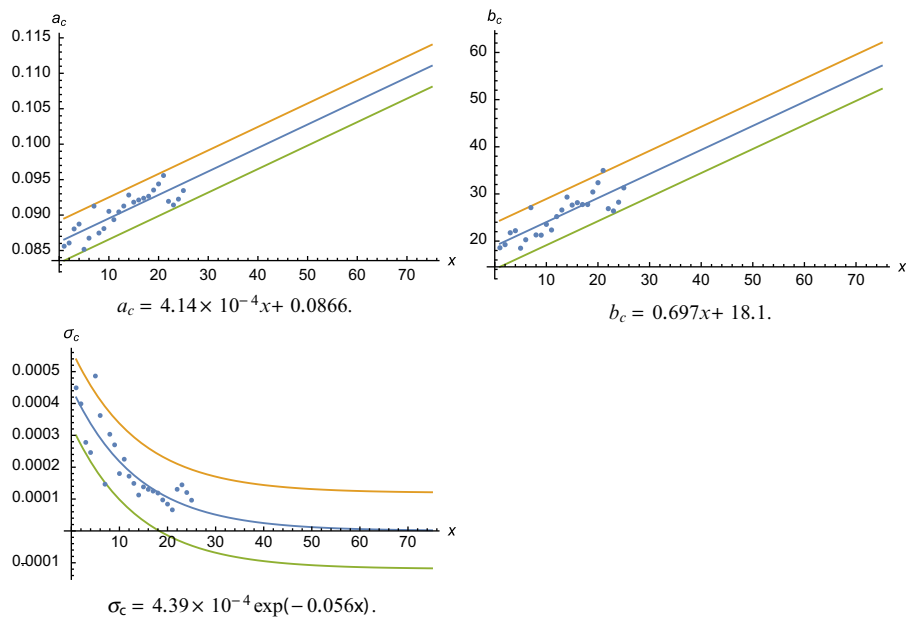
Figure 2. *Estimation of parameters by nonlinear regressions in IG-SEM of Denmark, Female. The x-axis represents $x = c - 1815$.*

2. To obtain the MPMF, we split the data into training and test data. For example, in $c' = 1890$, we split the mortality data into two parts: 20–60 years (training data: red dots) and 61–110 years (test data: black dots), and use the training data for modification (3.4).

3. In Figure 3, we will visually compare the two mortality curves with test data (black dots) for males and females, but only for $c' = 1890$. For other cohorts ($c' = 1850$ and 1870), we will only show the MSE between the predicted mortality function and the actual empirical mortality function in Table 3.

Table 2. *The (adjusted) coefficient of determination R^2 ($\overline{R}^2$) for nonlinear regression of each parameter with the 95%-prediction intervals (95%-PI). Although R^2 ($\overline{R}^2$) for Males is extremely small (the regression may not fit well), the MSE (Table 3) is not so bad. This is one of the advantages of our "modification".*

IG-SEM			
Females	$a_c = 4.14 \times 10^{-4}x + 0.0866$	$b_c = 0.697x + 18.12$	$\sigma_c = 4.39 \times 10^{-4} \exp(-0.056x)$
(95%-PI)	$\pm 2.89 \times 10^{-4}$	± 5.33	$\pm 1.11 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.817 (0.809)	0.787 (0.778)	0.827 (0.819)
Males	$a_c = 4.20 \times 10^{-5}x + 0.0826$	$b_c = -0.0211x + 18.19$	$\sigma_c = 1.23 \times 10^{-3} \exp(-0.010x)$
(95%-PI)	$\pm 2.97 \times 10^{-3}$	± 4.91	$\pm 1.19 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.086 (0.046)	0.055 (0.143)	0.156 (0.120)

Table 3. *MSE between MPMF and the empirical MF (test data) from Denmark data. Predictions for $c' = 1850, 1870$ are very good, and $c' = 1890$ is also admissible.*

MSE for MPMF	Females ID-SEM	Females IG-SEM	Males ID-SEM	Males IG-SEM
$c' = 1850$	3.75×10^{-8}	4.56×10^{-8}	4.23×10^{-9}	4.52×10^{-8}
$c' = 1870$	2.53×10^{-4}	7.42×10^{-4}	4.45×10^{-6}	5.57×10^{-5}
$c' = 1890$	3.84×10^{-4}	3.12×10^{-3}	8.78×10^{-5}	5.70×10^{-3}

Remark 4.2. We employed the simplest regression functions feasible to facilitate ease of use. For $\alpha_c, \beta_c, \kappa_c$, we used a negative increasing function of the form $-c_1 e^{-c_2 x} < 0$ because these values should be negative. Although γ_c should be positive, it may be justifiable to model it using a linear function, among other possible forms, given the available data. Occasionally, one can use the information criteria, for example, AIC or BIC, to select a regression function; it is also possible to use a time-series model to predict future parameters. However, any model has merits and demerits; therefore, we attempted it as simply as possible.

In this cohort (relatively long future prediction), the difference between ID-SEM and IG-SEM is more significant. Even a modified version in IG-SEM cannot predict well in males because of the parameter prediction for a_c and b_c. This is a successful example of ID-SEM with a change point T and more parameters than IG-SEM.

4.2 Norway

The second example is that of Norway. Similar to Denmark, we use the following mortality data from the $m = 25$ cohorts:

$$c_1 = 1826 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old;}$$
$$c_2 = 1827 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old;}$$
$$\vdots \qquad\qquad\qquad \vdots$$
$$c_{25} = 1850 \text{ birth cohort:} \quad 20 \text{ years old} - 110 \text{ years old.}$$

and assume that we are in 1961 (because we already have data for the 110-year-old in the 1840 cohort). Based on these data, we predicted the mortality functions of the 20-year-old for future cohorts $c' = 1860$, 1880, and 1900 for males and females. The prediction is after the 101 years for $c' = 1860$, 81 years for $c' = 1880$, and 61 years for $c' = 1900$. The results only for $c' = 1900$ are given in Figure 4.
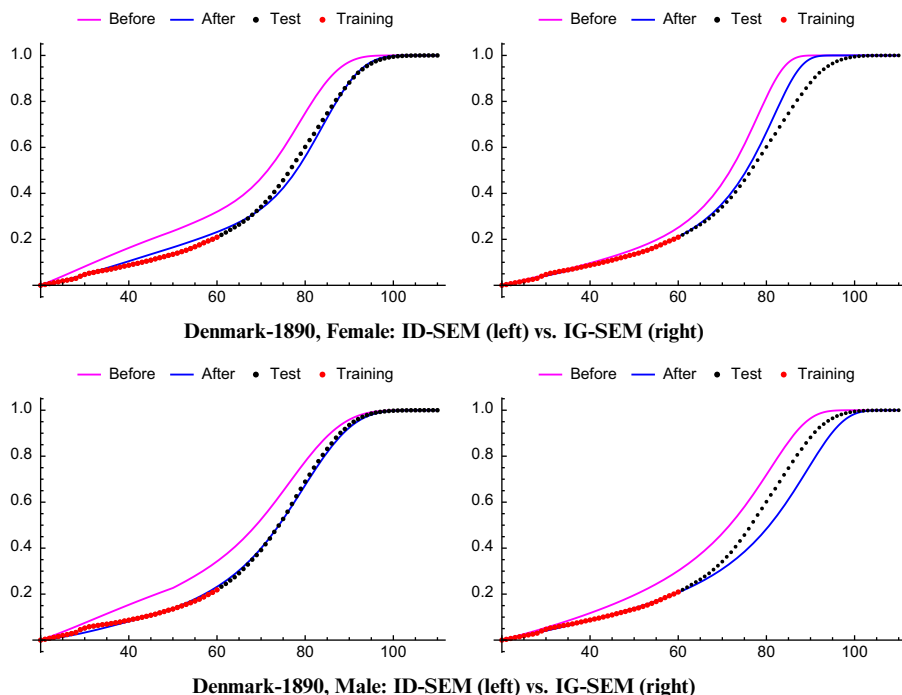
Figure 3. *Mortality functions by ID-SEM (left) and IG-SEM (right) for 1890 birth cohort in Denmark; females (top) and males (bottom). The magenta curve is before modification, and the blue one is the modified version. The prediction part is more than 60 years old.*

All other procedures were identical to those used in Denmark. We estimate parameters using non-linear regression and obtain PMFs before/after the modification. For these results, we only show the figures of PMF before/after changes for $c' = 1900$. For the others, we only show the nonlinear regression curves with the values of R^2 ($\overline{R}^2$) and their 95%-PI in Tables 4 and 5. Moreover, Table 6 lists the MSE of MPMFs.

Remark 4.3. In Denmark, ID-SEM is superior to IG-SEM. However, IG-SEM is effective in this example and occasionally outperforms ID-SEM. Because it is challenging to determine a suitable change-point parameter T in ID-SEM, IG-SEM, which has fewer parameters than ID-SEM, is also a good candidate for the prediction model of the mortality function.

In this example, IG-SEM is superior to ID-SEM in females but not males. Accordingly, it would be challenging to determine the SEM to predict and compute some quantities of interest. We should compute them both by ID-SEM and IG-SEM and compare the values objectively to make a decision

5. Advantages of SEM

5.1 Comparison with the classical model with cohort-effects

Shimizu *et al.* (2020) demonstrated that ID-SEM is superior to the classical Lee–Carter model. This section compares our SEM with the Renshaw–Haberman model (RHM), extending the Lee–Carter model, including cohort effects.

For comparison, we use the same data as in the previous section for Denmark and Norway. Moreover, for RHM, we used 20–110 years old of the 1911–1950 calendar years in Denmark and the same ages of 1921–1960 in Norway. We compared the modified mortality functions of the 1870 and 1890 birth cohorts using ID- and IG-SEM and the mortality functions of the RHM. The results are shown in Figure 5

Table 4. *The (adjusted) coefficient of determination R^2 ($\overline{R}^2$) for nonlinear regression of each parameter with the 95%-prediction intervals (95%-PI). For nonlinear exponential regression, we computed the R^2 ($\overline{R}^2$) by transforming it to the linear regression after taking the logarithm on both sides.*

ID-SEM

Females	$\alpha_c = -7.25 \exp(-0.011x)$	$\beta_c = -1.54 \exp(-0.013x)$
(95%-PI)	± 0.680	± 0.176
R^2 ($\overline{R}^2$)	0.682 (0.668)	0.635 (0.619)
Females	$\gamma_c = 5.54 \times 10^{-4}x + 0.131$	$\kappa_c = -2.03 \times 10^{-3} \exp(-0.020x)$
(95%-PI)	$\pm 4.96 \times 10^{-3}$	$\pm 3.07 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.730 (0.718)	0.692 (0.679)
Males	$\alpha_c = -7.41 \exp(-0.013x)$	$\beta_c = -2.16 \exp(-0.015x)$
(95%-PI)	± 0.621	± 0.222
R^2 ($\overline{R}^2$)	0.782 (0.772)	0.761 (0.751)
Males	$\gamma_c = 7.13 \times 10^{-4}x + 0.127$	$\kappa_c = -1.79 \times 10^{-3} \exp(-0.0259x)$
(95%-PI)	$\pm 5.07 \times 10^{-3}$	$\pm 2.50 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.811 (0.803)	0.759 (0.749)

Table 5. *The (adjusted) coefficient of determination R^2 ($\overline{R}^2$) for nonliner regression of each parameter with the 95%-prediction intervals (95%-PI).*
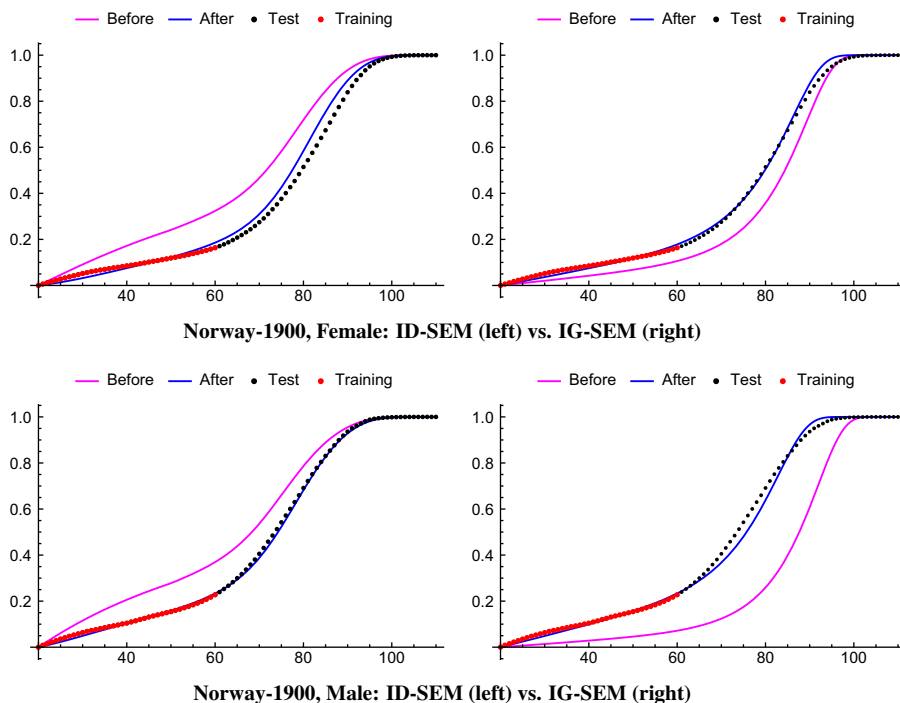
IG-SEM

Females	$a_c = 3.31 \times 10^{-4}x + 0.0862$	$b_c = 0.510x + 18.93$	$\sigma_c = 4.50 \times 10^{-4} \exp(-0.073x)$
(95%-PI)	$\pm 2.97 \times 10^{-3}$	± 4.90	$\pm 1.19 \times 10^{-4}$
R^2 ($\overline{R}^2$)	0.729 (0.717)	0.701 (0.688)	0.737 (0.726)
Males	$a_c = 3.69 \times 10^{-4}x + 0.0889$	$b_c = 0.787x + 24.04$	$\sigma_c = 3.42^{-4} \exp(-0.090x)$
(95%-PI)	$\pm 3.48 \times 10^{-3}$	± 9.18	$\pm 8.00 \times 10^{-5}$
R^2 ($\overline{R}^2$)	0.708 (0.695)	0.614 (0.597)	0.704 (0.691)

Table 6. *MSE between MPMF and the empirical MF (test data) for Norway data. After our modification, the predictions become very good in any case.*

MSE for MPMF	Females		Males	
	ID-SEM	IG-SEM	ID-SEM	IG-SEM
$c' = 1860$	3.07×10^{-8}	5.48×10^{-8}	1.46×10^{-9}	1.80×10^{-7}
$c' = 1880$	1.46×10^{-4}	8.57×10^{-5}	3.66×10^{-5}	1.01×10^{-4}
$c' = 1990$	1.91×10^{-3}	2.51×10^{-4}	9.20×10^{-5}	9.41×10^{-4}

along with their MSEs. The results demonstrate that the differences in prediction errors are similar, but ID-SEM is often superior to RHM at senior ages.

Remark 5.1. Although we used the CBD model, for example, Cairns *et al.* (2006a, 2008), as a candidate cohort model, it was unsuitable for long-term prediction. Therefore, these results were excluded from this study.

Figure 4. *Mortality functions by ID-SEM (left) and IG-SEM (right) for 1900 birth cohort in Norway; females (top) and males (bottom). The magenta curve is before modification, and the blue one is the modified version. The prediction part is more than 60 years old.*

5.2 Reducing statistical errors

One of the advantages of the proposed SEM approach is the statistical estimation of the actuarial quantities. Consider, for example, the single premium of all life insurance at age x, say A_x. It is written as follows:

$$A_x := \sum_{k=1}^{\infty} v^k{}_{k-1|}q_{x+k-1} \quad \text{(Actuarial notation)}$$

$$= \sum_{k=1}^{\infty} v^k \frac{q_c(x+k) - q_c(x+k-1)}{1 - q_c(x)} \quad \text{(SEM notation)}$$

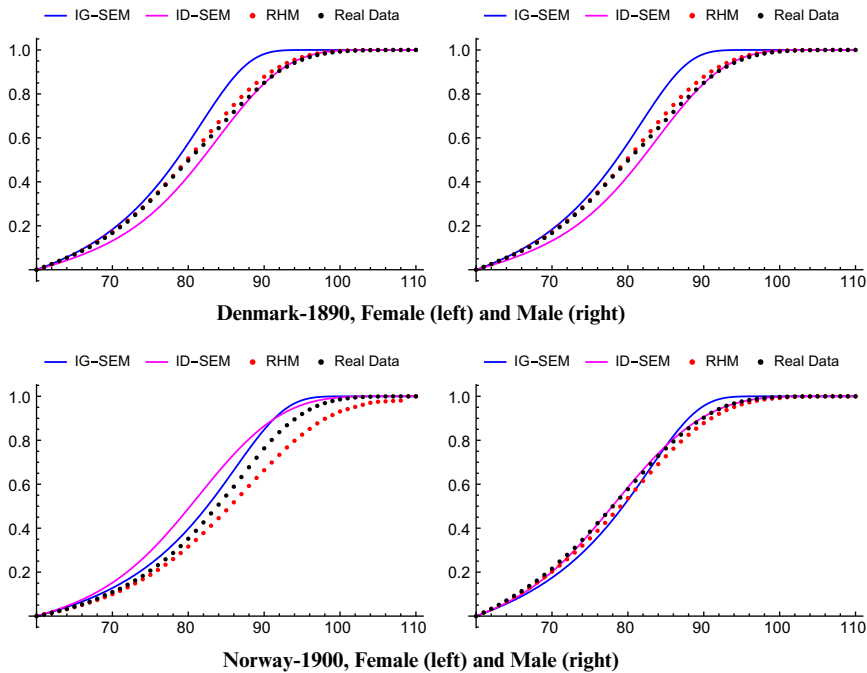$$= \sum_{k=1}^{\infty} v^k [q_c(k|x) - q_c(k-1|x)] \quad \text{(Conditional version)},$$

where $v \in (0, 1)$ is the discount factor. If we use the Lee–Carter model, then it is written as

$$A_x = \sum_{k=1}^{\infty} v^k \left[1 - \exp\left(-m_{x+k-1,t}(\alpha_{x+k}, \beta_{x+k}) \right) \right],$$

where $m_{x,t}$ is the (crude) mortality parameterized by

$$m_{x,t}(\alpha_x, \beta_x) = \exp\left(\alpha_x + \beta_x \kappa_t + \epsilon_{x,t} \right),$$

with parameters α_x, β_x estimated based on the predicted values of κ_t, which are generated using a time series model that includes some unknown parameters, and $\epsilon_{x,t}$ is a noise process. Here, we must estimate

Figure 5. *Modified mortality functions by IG-SEM (blue); ID-SEM (magenta), and RHM (red dots) with a table of their MSEs for 1890 cohort of Denmark (top) and 1900 cohort of Norway (bottom). The black dots are the actual data that should be predicted. The results for males suggest that ID-SEM is superior, whereas those for females suggest that IG-SEM and RH may be better, depending on the case. Ultimately, which model is most suitable depends heavily on the data.*

numerous parameters $\{(\alpha_y, \beta_y)\}_{y=x,x+1,\dots}$ and those in κ_t, which can increase the statistical error of A_x. However, if we use SEM, cohort-wise computation

$$A_x = \sum_{k=1}^{\infty} v^k [q_c(k, \vartheta_c|x) - q_c(k-1, \vartheta_c|x)],$$

requires only one parameter estimation for ϑ_c because ϑ_c is independent of $k = 1, 2, \dots$. This can make the statistical error less than that of classical mortality models.

5.3 Sensitivity analysis

As shown in the previous section, most actuarial quantities are written in the functionals of the mortality function $q_c(t, \vartheta_c)$, which are often rewritten in terms of the conditional mortality function $q_c(t, \vartheta_c|S)$, with

a few unknown parameters ϑ_c. This situation is suitable for *sensitivity analysis* concerning parameter changes.

Consider an actuarial quantity for age x and cohort c represented by a Stieljes-type integral form such as

$$H(\vartheta) := \int_0^\infty q_c(t, \vartheta | x) \, \mathrm{d}h(t), \quad \vartheta \in \Theta,$$

where h denotes a measurable function of $[0, \infty)$, The integral sign implies that $\int_0^\infty := \int_{[0,\infty)}$. We suppose the exchangeability of $\int_0^\infty$ and differentiation ∂_ϑ as far as we need

$$\partial_\vartheta H(\vartheta) = \int_0^\infty \partial_\vartheta q_c(t, \vartheta | x) \, \mathrm{d}h(t) < \infty, \quad \vartheta \in \Theta,$$

which is continuous in ϑ.

Most actuarial quantities are written in this form (see Shimizu *et al.* (2020)). For example, A_x, the single premium of all life insurance at age x, is given by

$$h(t) = \sum_{k=1}^\infty v^k \left(\mathbf{1}_{\{t \geq k\}} - \mathbf{1}_{\{t \geq k-1\}} \right), \quad t \geq 0,$$

where $v \in (0, 1)$, Moreover, for the immediate payment version:

$$\bar{A} = \int_0^\infty v^t \frac{\partial_t q_c(x+t)}{1 - q_c(x)} \, \mathrm{d}t,$$

is given by $H(\vartheta)$ with

$$h(t) = -v^t.$$

It follows from integration by parts that

$$
\begin{aligned}
H(\vartheta) &= \int_0^\infty \frac{q_c(x+t, \vartheta) - q(x, \vartheta)}{1 - q_c(x, \vartheta)} (-v^t)' \mathrm{d}t \\
&= \left[-v^t \frac{q_c(x+t, \vartheta) - q_c(x, \vartheta)}{1 - q_c(x, \vartheta)} \right]_{t=0}^\infty + \int_0^\infty v^t \frac{\partial_t q_c(x+t, \vartheta)}{1 - q_c(x, \vartheta)} \, \mathrm{d}t \\
&= \int_0^\infty v^t \frac{\partial_t q_c(x+t, \vartheta)}{1 - q_c(x, \vartheta)} \, \mathrm{d}t = \bar{A}_x.
\end{aligned}
$$

We are interested in the difference $H(\vartheta) - H(\vartheta_c)$ for different values of parameters ϑ and ϑ_c. By Taylor's formula,

$$H(\vartheta) - H(\vartheta_c) = \int_0^\infty \partial_\vartheta q_c(t, \vartheta_c | x) \, \mathrm{d}h(t) \cdot (\vartheta - \vartheta_c) + o(\vartheta - \vartheta_c).$$

Integral $\int_0^\infty \partial_\vartheta q_c(t, \vartheta_c | x) \, \mathrm{d}h(t)$ can be evaluated via direct computation. For instance, we have the following inequality:

Lemma 5.2. *For the mortality function of IG-SEM, $q_c^{IG}(t, \vartheta)$ with $\vartheta = (a, b, \sigma)$, we obtain the following estimates:*

$$\left| \partial_a q_c^{IG}(t, \vartheta) \right| \leq 2t e^{at} \left(1 + \frac{\sigma}{\Lambda_\vartheta(t) + x_c} \right) \phi_{x_c, x_c/\sigma}(\Lambda_\vartheta(t));$$

$$\left| \partial_b q_c^{IG}(t, \vartheta) \right| \leq 2t \left(1 + \frac{\sigma}{\Lambda_\vartheta(t) + x_c} \right) \phi_{x_c, x_c/\sigma}(\Lambda_\vartheta(t));$$

$$\left| \partial_\sigma q_c^{IG}(t, \vartheta) \right| \leq \Lambda_\vartheta(t) \left(\frac{1}{\sigma} + \frac{2}{\Lambda_\vartheta(t) + x_c} \right) \phi_{x_c, x_c/\sigma}(\Lambda_\vartheta(t)),$$

where x_c is the initial survival energy for cohort c and $\phi_{u,v}(x)$ is the probability density function of the normal distribution with mean u and variance v.

Proof. Note that

$$q_c^{IG}(t, \vartheta) = \Phi\left(\sqrt{\frac{\sigma}{x_c}}(\Lambda_\vartheta(t) - x_c)\right) - e^{2\sigma\Lambda_\vartheta(t)}\Phi\left(-\sqrt{\frac{\sigma}{x_c}}(\Lambda_\vartheta(t) + x_c)\right),$$

where $\Phi(x) = (2\pi)^{-1/2}\int_{-\infty}^x e^{-z^2/2}\,dz$.

$$\begin{aligned}
\partial_a q_c^{IG}(t, \vartheta) &= te^{at}\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) - 2\sigma te^{at}e^{2\sigma\Lambda_\vartheta(t)}\Phi\left(-\sqrt{\frac{\sigma}{x_c}}(\Lambda_\vartheta(t) + x_c)\right) \\
&\quad + e^{2\sigma\Lambda_\vartheta(t)}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t))te^{at} \\
&= 2te^{at}\left[\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) - \sigma e^{2\sigma\Lambda_\vartheta(t)}\Phi\left(-\sqrt{\frac{\sigma}{x_c}}(\Lambda_\vartheta(t) + x_c)\right)\right].
\end{aligned}$$

In the final equality, we used

$$e^{2\sigma\Lambda_\vartheta(t)}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t)) = \phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)). \tag{5.1}$$

We use an inequality for the "error function" such that for $x > 0$,

$$\Phi(-x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz \le \frac{1}{\sqrt{2\pi}x}e^{-\frac{x^2}{2}},$$

to obtain

$$\begin{aligned}
\left|\partial_a q_c^{IG}(t, \vartheta)\right| &\le 2te^{at}\left[\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) + \sigma e^{2\sigma\Lambda_\vartheta(t)}\Phi\left(-\sqrt{\frac{\sigma}{x_c}}(\Lambda_\vartheta(t) + x_c)\right)\right] \\
&\le 2te^{at}\left[\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) + \frac{\sigma}{\Lambda_\vartheta(t) + x_c}e^{2\sigma\Lambda_\vartheta(t)}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t))\right] \\
&= 2te^{at}\left(1 + \frac{\sigma}{\Lambda_\vartheta(t) + x_c}\right)\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)).
\end{aligned}$$

We used equality (5.1) in the last equality. The estimate of the partial derivative $\partial_b q_c^{IG}$ is slightly similar and omitted.

For $\partial_\sigma q_c^{IG}$, it follows from (5.1) that:

$$\begin{aligned}
\partial_\sigma q_c^{IG}(t, \vartheta) &= \frac{1}{2\sqrt{\sigma x_c}}(\Lambda_\vartheta(t) - x)\cdot\sqrt{\frac{x_c}{\sigma}}\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) - 2\Lambda_\vartheta(t)e^{2\sigma\Lambda_\vartheta(t)}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t)) \\
&\quad + \frac{1}{2\sqrt{\sigma x_c}}(\Lambda_\vartheta(t) + x)\cdot\sqrt{\frac{x_c}{\sigma}}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t)) \\
&= \frac{1}{\sigma}\Lambda_\vartheta(t)\phi_{x_c,x_c/\sigma}(\Lambda_\vartheta(t)) - 2\Lambda_\vartheta(t)e^{2\sigma\Lambda_\vartheta(t)}\phi_{-x_c,x_c/\sigma}(\Lambda_\vartheta(t)).
\end{aligned}$$

Hence, the same argument as above is available and the proof ends. □

Corollary 5.3. *Under the same model as in Lemma 5.2, assume that:*

$$\sup_{\vartheta\in\Theta}\left|\int_0^\infty te^{at}\phi_{0,1}(\Lambda_\vartheta(t))\,dh(t)\right| < \infty.$$

Subsequently, it follows that

$$\sup_{\vartheta\in\Theta}\left|\int_0^\infty \partial_\vartheta q_c(t, \vartheta|x)\,dh(t)\right| < \infty.$$

For our LSE $\widehat{\vartheta}_c$ of ϑ_c given in Theorem 3.2 in Shimizu *et al.* (2020) (see also the erratum Shimizu, 2022) and the sample size n_c required to obtain the estimator, we have, by the delta method in statistics, that

$$\sqrt{n_c}\left(H(\widehat{\vartheta}_c) - H(\vartheta_c)\right) = \int_0^\infty \partial_\vartheta q_c(t, \vartheta_c|x) \, dh(t) \cdot \sqrt{n_c}(\widehat{\vartheta}_c - \vartheta_c) + o_p(1)$$
$$\to^d N_p(0, \Sigma_{c,x}), \quad n_c \to \infty,$$

where the asymptotic variance $\Sigma_{c,x}$ can be estimated using the estimators of R_d, Q_d, Σ in Theorem 3.2 in Shimizu *et al.* (2020) (with Shimizu, 2022), and the plug-in estimator $\int_0^\infty \partial_\vartheta q_c(t, \widehat{\vartheta}_c|x) \, dh(t)$. This yields the confidence interval $H(\vartheta_c)$:

$$\mathbb{P}\left(H(\vartheta_c) \in \left[H(\widehat{\vartheta}_c) - z_{\alpha/2}\frac{\widehat{\Sigma}_{x,c}}{\sqrt{n_c}}, \ H(\widehat{\vartheta}_c) + z_{\alpha/2}\frac{\widehat{\Sigma}_{x,c}}{\sqrt{n_c}}\right]\right) \approx 1 - \alpha,$$

where z_α is the upper α-percentile of the standard normal distribution and $\widehat{\Sigma}_{c,x}$ is an estimator of the asymptotic variance $\Sigma_{c,x}$.

6. Conclusions

We proposed two types of parametric families for SEMs: ID-SEM and IG-SEM, which provide accurate cohort-wise PMFs. Using the (prediction) confidence intervals for unknown parameters, we can modify the MPMF to fit existing data in a manner consistent with LSE (refer to Remark 3.1).

SEM is a viable candidate for alternative modeling of mortality prediction. We illustrated that both SEMs had high potential for long-term mortality prediction and were superior to the classical model, possibly with cohort effects, for example, LC, RH, and CBD models. Moreover, SEM has numerous theoretical advantages: notational understanding for nonactuarial people, reduced estimation error owing to fewer parameters, and usefulness for sensitivity analysis.

For further information regarding SEM, such as graphs and other topics, please refer to the supplementary article by Shimizu *et al.* (2023).

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 J. Financ. Math.*, **3**, 639–666.

Baukai, B. (1990) Explicit expression for the distribution of the supremum of Brownian motion with a change point. *Commun. Stat. Theory Methods*, **19**(1), 31–40.

Biffis, E. (2005) Affine processes for dynamic mortality and actuarial valuation. *Insur. Math. Econ.*, **37**, 443–468.

Biffis, E., Denuit, M. and Devolder, P. (2010) Stochastic mortality under measure changes. *Scand. Actuar. J.*, (4), 284–311.

Blackburn, C. and Sherris, M. (2013) Consistent dynamic affine mortality models for longevity risk applications. *Insur. Math. Econom.*, **53**, 64–73.

Cairns, A.J.G., Blake, D. and Dowd, K. (2006a) A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *J. Risk Insur.*, **73**, 687–718.

Cairns, A.J.G., Blake, D. and Dowd, K. (2006b) Pricing death: Framework for the valuation and securitization of mortality risk. *ASTIN Bull.*, **36**(1), 79–120.

Cairns, A.J.G., Blake, D. and Dowd, K. (2008) Modeling and management of mortality risk: A review. *Scand. Actuarial J.*, (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 Am. Actuarial J.*, **13**(1), 1–35.

Chen, H. and Cox, S.H. (2009) Modeling mortality with jumps: Applications to mortality securitization. *J. Risk Insurance*, **76**(3), 727–751.

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

For SEM data processing: https://www.shimizu.sci.waseda.ac.jp/smzlab/files/To_semData.r.

Hainaut, D. and Devolder, P. (2008) Mortality modeling with Lévy processes. *Insur. Math. Econ.*, **42**, 409–418.

Hao, X., Li, X. and Shimizu, Y. (2013) Finite-time survival probability and credit default swap pricing in geometric Lévy markets. *Insur. Math. Econ.* **53**, 14–23.

Human Mortality Database: https://www.mortality.org/.

Ito, R. and Shimizu, Y. (2019) Cohort-wise mortality prediction under the survival energy hypothesis (in Japanese). *J. Jpn. Assoc. Risk Insur. Pensions (JARIP)*, **6**, 17–30.

Jarrow, R.A. and Turnbull, S. (1995) Pricing derivatives on financial securities subject to credit risk. *J. Finance*, **50**, 53–85.

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

Lee, R.D. and Carter, L. (1992) Modeling and forecasting the time series of US mortality. *J. Am. Stat. Assoc.*, **87**, 659–671.

Leng, X. and Peng, L. (2016) Inference pitfalls in the Lee-Carter model for forecasting mortality. *Insur. Math. Econ.*, **70**, 58–65.

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

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

Merton, R. (1976) Option pricing when underlying stock returns are discontinuous. *J. Financ. Econ.*, **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.

Olovieri, A. (2001) Uncertainty in mortality projections: An actuarial perspective. *Insur. Math. Econ.*, **29**, 231–245.

Pitacco, E. (2004) Survival models in a dynamic context: A survey. *Insur. Math. Econ.*, **35**, 279–298.

Renshaw, A.E. and Haberman, S. (2003) Lee-Carter mortality forecasting with age-specific enhancement. *Insur. Math. Econ.*, **33**, 255–272.

Renshaw, A. and Haberman, S. (2006) A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insur. Math. Econ.*, **38**, 556–570.

Schoutens, W. and Cariboni, J. (2009) *Lévy Processes in Credit Risk*. Oxford (United Kingdom): John Wiley & Sons.

Shimizu, Y., Minami, Y. and Ito, R. (2020) Why does a human die? A structural approach to cohort-wise mortality prediction under the survival energy hypothesis. *ASTIN Bull.*, **51**(1), 191–219.

Shimizu, Y. (2022) Erratum to: "Why does a human die? A structural approach to cohort-wise mortality prediction under Survival Energy Hypothesis." [ASTIN Bull., 2021; 51(1) 191–219.], https://www.shimizu.sci.waseda.ac.jp/files/Erratum-ASTIN51-1.pdf.

Shimizu, Y., Shirai, K., Kojima, Y., Mitsuda, D. and Inoue, M. (2023) A new survival energy model and SEM project. SSRN:4127900: http://dx.doi.org/10.2139/ssrn.4127900.

Shirai, K. and Shimizu, Y. (2022) *On the prediction of full life expectancy via Survival Energy Models*. Master Thesis, Graduate School of Fundamental Science and Engineering, Waseda University (in Japanese); *submitted*.

van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Ye, Z. and Chen, N. (2014) The inverse Gaussian process is a degradation model. *Technometrics*, **56**(3), 302–311.