A Semiparametric Method for Assessing Life Expectancy Evaluations

Hong Beng Lim[a*] and Nariankadu D. Shyamalkumar[a]

[a] *Department of Statistics and Actuarial Science, The University of Iowa; Iowa City, IA, U.S.A.*

[*] Corresponding author. Email: *hongbeng-lim@uiowa.edu*

Abstract: In the life settlements industry, Life Expectancy (LE) providers are firms which conduct health underwriting towards predicting the future mortality of an insured. Multiple stakeholders are interested in assessing the quality of their evaluations. There has been some recent interest in better alternatives to the traditional metric for this quality, the A/E ratio: the ratio of *actual* to *expected* number of deaths. One such alternative is the *Implied Difference in Life Expectancies* (IDLE) metric proposed by Bauer et al. (2018). Its design largely retains the simplicity of the A/E ratio while being informative, unlike the A/E ratio, throughout the life of a policy block. Even though the IDLE is a significant improvement over the A/E ratio, it turns out that the IDLE is sensitive to departures from a key assumption, which motivates our development of a more robust metric. We propose the use of a Cox proportional hazards model for this purpose with the covariates derived from the information contained in the LE provider's report. We present evidence that the deviation of the future life expectation evaluations from those derived using this model is a worthy alternative to both the A/E ratio and the IDLE.

1. Introduction

A life insurance policyholder may seek a feasible way out of a policy for reasons such as no longer having dependents or the ability to pay premiums. In such cases, aside from lapsing or surrendering the policy, certain policyholders may benefit from the liquidity provided by life settlements. These are transactions in which the purchaser, in exchange for being named the beneficiary on the policy, offers the policyholder a lump sum (which is above the surrender value but less than the death benefit), besides assuming the obligation of the future premiums.

The value proposition of life settlements hinges upon the fact that the cash surrender values are determined based on the health of the insured at or near policy inception, whereas the current health of the insured may indicate a different valuation. The actuarial value of a life policy is directly related to the insured's life expectancy: the shorter the insured lives, the less premiums paid and the earlier the death benefit is received, hence the policy is worth more. Therefore, if the insured's health has deteriorated since issue, the policy may be worth much more than the surrender value. In such cases a life settlement can help unlock part of this excess value. Indeed, studies have estimated that on those policies that were settled, the offer value was on average about four times that of the surrender value (Januário and Naik, 2016; Doherty and Singer, 2003). Notably, this is in spite of the fact that the life settlements markets remain quite inefficient; a substantial portion of the expected present value goes towards transaction costs and profit margins (Deloitte, 2005).

The selling of life policies first became popular during the AIDS crisis of the 1980s. Then *viatical settlements* offered AIDS patients much needed cash to seek medical treatment. As the advent of antiretroviral drugs in the late 1990s radically improved patients' life expectancies, the market focus gradually shifted to senior policyholders with below-average health, giving rise to the life settlement markets of today. The secondary market is where the policies are first purchased, whereas investors trade such policies in the tertiary market. At about 9 billion (USD) death benefit worth of insurance policies traded per year, the tertiary market is currently more active than the secondary market, which has about 2.5 billion (USD) face value transacted per year (Braun and Xu, 2017). With an aging US population, increased streamlining of the transaction process as well as efforts to market the life settlement option to the general public, these markets are poised to grow in the coming years (Gibson, 2019).

The life settlement transaction is usually initiated by a life settlement company which is an intermediary specializing in the purchase of life policies and their securitization. Securitization results in providing access

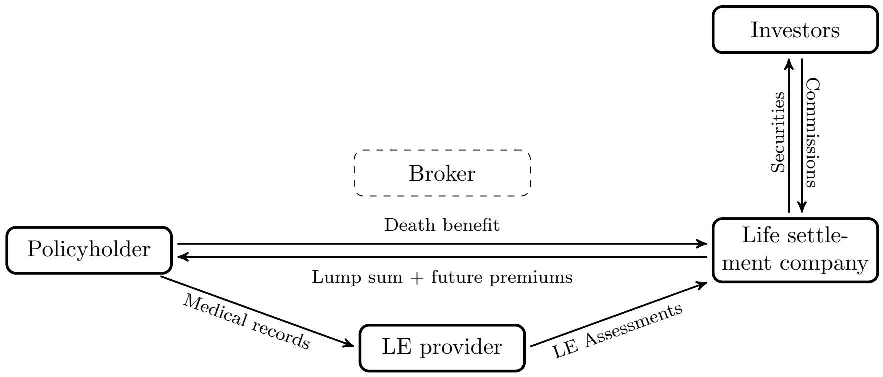


Figure 1: Simplified market dynamics of the US life settlement markets.

to investors typically through the creation of specialized funds (Braun et al., 2015) or life settlement-backed securities (Stone and Zissu, 2006). For policyholder initiated transactions, a life settlement broker may serve as another intermediary, and as part of the service collect bids from a few life settlement companies. A key input in determining the bid price is the life expectancy (LE) assessment provided by medical underwriters known as LE providers. These firms provide an assessment of the insured's future mortality based upon a review of the insured's medical records; this is used to calculate the actuarial value of the policy. A simplified diagrammatic representation of this process is depicted in Figure 1.

The life settlement market attracts investors as the risk entailed in its cashflows are uncorrelated with the typical financial market risks (Braun et al., 2012), thus making them a useful tool for portfolio diversification strategies (Davó et al., 2013). Nevertheless, such investments still contain largely non-diversifiable risk such as the longevity risk and the valuation risk introduced by the LE assessment process (A. M. Best, 2016; Braun et al., 2012). In general, LE providers assess the future mortality of insureds by assigning debits and credits to each of his/her medical conditions in order to produce a mortality multiplier. This multiplier is then applied to an appropriate baseline mortality table to arrive at an estimate of the future mortality distribution for the insured. Estimates of all measures of future mortality, such as the life expectancy, are derived from this distribution. In the case of insureds with complex health histories, however, clinical judgement may instead be used, resulting in a more customized LE assessment (Braun and Xu, 2017). The quality of LE providers' assessments hence depends not only on the LE providers' modeling assumptions, but very much also on the care taken in constructing an up to date health history of the insured.

This quality is of interest to various external stakeholders. Investors mainly need to protect themselves against positive bias in mortality predictions which inflate the offer price and lead to unexpected lower portfolio returns. Conversely, these higher offer prices potentially benefit life settlement companies and brokers who are compensated by commissions on the consummated transactions. As these intermediaries are also the clients of the LE providers, it is no surprise that LE assessments have tended to be aggressive (short) in the early years of the market (European Life Settlement Association, 2013; Braun et al., 2015). To ensure the sustainability of the life settlements market, it is crucial to balance these opposing interests: that is, for LE providers' estimates to be as accurate as possible. Currently regulators are only monitoring the quality of LE providers in the state of Florida (The Florida Legislature, 2019), but more states may begin to institute similar requirements as the market expands.

The predominant metric used to assess the quality of LE assessments is the A/E (actual-to-expected) ratio of deaths. This is the ratio of the number of deaths occurring within a period of time over the number of expected deaths within the period. Note that the latter quantity equals the sum of the probabilities that individual insureds die within the period, and hence can be derived from the LE assessments. While high-quality estimates is expected to result in an A/E ratio close to 1, the reverse implication is problematic, especially for seasoned portfolios. This is so as mortality pulls the A/E ratio to 1 regardless of systematically biased LE assessments (see e.g. Bauer et al. (2018)).

As an alternative, Bauer et al. (2018) proposes the *Implied Difference in Life Expectancies* (IDLE) metric, which infers the degree to which the life expectancy assessments are systematically misstated based on deviations of the term life expectancies from its sample analog. While more informative and consistently interpretive than the A/E ratio, this methodology relies on the assumption that the underlying death probabilities are a multiple of those in the reference mortality table employed by their methodology. As is alluded to in Bauer et al. (2018), and as we illustrate in the paper, under departures from this assumption the metric is materially biased. This motivates our search to develop a metric which is similarly interpretive as the IDLE, but which is more robust.

Such a robust metric will be particularly useful to external stakeholders in evaluating the quality of a LE provider's assessments as they do not necessarily have access to the LE provider's proprietary mortality tables. For example, a life settlement company/broker may wish to compare LE providers, prior to choosing the service of one. Independent actuaries could also be commissioned to assess the quality of an LE provider, who may not wish to disclose their in-house mortality tables to the investigators (Actuarial Standards Board, 2013). Notably, there has been some reluctance about making A/E reports publicly available, since one may be able to infer the proprietary table from such reports (Braun and Xu, 2017). A robust metric would facilitate informative reports without compromising the LE providers' core competency.

In this paper, we will show that, if the LE provider's assessments are sufficiently informative about mortality of the insured, that covariates useful for semiparametric survival models can be derived from these assessments. We propose as a metric the average difference of life expectancies estimated from such models from that of the LE provider's assessments, which we will call the *Estimated Difference in Life Expectancies* (EDLE). We will show that this metric is significantly less biased than the IDLE under a misspecified reference table, and performs comparably when the reference table is correctly specified, hence demonstrating that it is a robust alternative to the IDLE and the A/E ratio.

1.1. Definitions and Notation

We use conventional actuarial notation ${}_tp_x$ (resp., ${}_tq_x$) to denote the probability that a person aged x survives (resp., dies within) the next t years. When $t = 1$, it is left out as in p_x, and q_x. Standard mortality tables list q_x's for a range of integer values of x. The mortality tables employed in our study are the versions for underwriting age equal to the exact age on the birthday nearest to the underwriting date (*i.e.* ANB tables). In our simulations, we subscribe to the constant force fractional age assumption. In other words, the constant force of mortality assumption holds if, for x an integer and $0 < s < 1$,

$$\mu_{x+s} = -\log p_x.$$

The device of mortality multipliers (sometimes referred to as mortality ratings) allows one to adjust a baseline mortality curve to account for heterogeneity. We use the superscript "(λ)" to denote that the underlying table is the baseline table adjusted by using λ as a multiplier. Typically, annual death probabilities are scaled by λ, that is $q_x^{(\lambda)} = \min(1, \lambda \cdot q_x)$ to be precise. We will call this the *constant multiple* formulation. For mathematical convenience, we instead adopt the *proportional hazards* formulation for mortality multipliers. Under this formulation we have $p_x^{(\lambda)} = (p_x)^\lambda$, and equivalently, under the constant force assumption, we have $\mu_x^{(\lambda)} = \lambda \cdot \mu_x$. Incidentally, this formulation as the name suggests is coherent with the Cox proportional hazards model. Importantly, when q_x is small, the $q_x^{(\lambda)}$ in this formulation is approximately equal to $\lambda \cdot q_x$.

By T_x, a random variable, we denote the time until death of a person aged x. Its expectation is denoted by $\mathring{e}_x$, the complete life expectancy for a person aged x. Similarly, we denote the t-year term complete life expectancy of a person aged x by $\mathring{e}_{x:\overline{t}|}$. Note that,

$$\mathring{e}_x = \mathbb{E}[T_x] = \int_0^\infty {}_tp_x dt \quad \text{and} \quad \mathring{e}_{x:\overline{t}|} = \mathbb{E}[\min(T_x, t)] = \int_0^t {}_sp_x ds. \tag{1}$$

To explicitly refer to the underlying mortality table and the multiplier, we will resort to using modifications such as $\mathring{e}_x^{Table,(\lambda)}$. When not otherwise qualified, all life expectancies refer to their complete versions.

2. Sensitivity Analysis of the IDLE

In this section we begin by lending some insight into fact that the IDLE metric of Bauer et al. (2018) may be biased when the shape of the baseline reference table does not match that of the true underlying mortality. Moreover, through a simulation study we study the extent to which the IDLE may be biased.

In this paper, we will assume to be working in the below described setting, which is motivated from the setting of the US life settlements market. Consider a policy block consisting of n insureds underwritten by a single LE provider that is to serve as the basis for evaluating the quality of this provider's LE assessments. We will assume that for this purpose the LE provider's underwriting report for each insured is available, and this contains an estimated life expectancy, $e(i)$, for the i-th insured. This equates to that corresponding to the included reference table and the stated mortality multiplier. The t-year term complete life expectancy corresponding to the i-th insured is denoted by $e_{\overline{t}|}(i)$, which is derived using the reference table and the mortality multiplier. We note that some companies are reluctant to reveal their proprietary mortality tables, and resort to instead using a reference table and a mortality multiplier for conveying their mortality prediction. If no table is specified for a mortality quantity, we understand it to be derived using the reference table, *e.g.* as in $\mathring{e}_x^{(\lambda)}$.

2.1. The IDLE

Consider a portfolio consisting of n settled life policies issued to n policyholders. For each $i = 1, \ldots, n$, let x_i denote the underwriting age of the i-th policyholder at the time the policy on his/her life was settled. We assume that their individual forces of mortality are proportional to that of some baseline, say μ_x, and denote by λ_i their individual mortality multipliers. We will by $T(i)$ denote the complete future lifetime random variable corresponding to the i-th insured. Note that, $\mathbb{E}[T(i)]$ may not equal $e(i)$. Further, let t_i be the time since the life policy on the i-th insured was settled. This introduces right-censoring in the sense that, we observe $T(i) \wedge t_i := \min(T(i), t_i)$, and $\mathbf{1}(T(i) < t_i)$, but not $T(i)$.

In Bauer et al. (2018), the authors posit that a life product independent measure of the quality of LE assessments is given by

$$\frac{1}{n} \sum_{i=1}^{n} (T(i) - e(i)), \tag{2}$$

which (by subscribing to a law of large numbers) can be seen to converge to zero if there is no material bias in the life expectancy assessments. This measure is referred to in Bauer et al. (2018) as the Difference in Life Expectancies (DLE), But since for any recent or open policy block a significant number of policies will be active, computing the DLE would, due to censoring, require restricting to inactive policies. To avoid the significant loss of sample information this would entail, Bauer et al. (2018) instead suggest the use of a censored version of this quantity given by

$$\frac{1}{n} \sum_{i=1}^{n} (T(i) \wedge t_i - e_{\overline{t_i}|}(i)), \tag{3}$$

which they call the Difference in Temporary Life Expectancies (DTLE). While the DTLE approaches the DLE as t_i's increase for each $i = 1, \ldots, n$, for an active block of policies the difference may be significant. Moreover, the DTLE across policy blocks are no longer comparable and hence cannot serve as another easy-to-interpret measure of quality of the LE assessments.

The IDLE in a sense is an extrapolation from the DTLE to the DLE. There are two versions of the IDLE, namely the *absolute* and the *relative* versions. Towards defining the absolute version, let $\lambda_{ai}(d)$ denote the mortality multiplier such that

$$\mathring{e}_{x_i}^{(\lambda_{ai}(d))} = e(i) + d \tag{4}$$

Then the absolute IDLE, which we denote by $\hat{d}$, is the discrepancy d such that

$$\frac{1}{n}\sum_{i=1}^{n} \mathring{e}^{(\lambda_{ai}(d))}_{x_i:\overline{t_i}|} = \frac{1}{n}\sum_{i=1}^{n}(T(i) \wedge t_i). \tag{5}$$

That is, $\hat{d}$ is the uniform discrepancy in the complete life expectancy which, via the modified mortality multipliers $\lambda_{ai}(d)$ leads to zero discrepancy between the average term life expectancy and its empirical analog. A *relative* IDLE can also be defined: let $\lambda_{ri}(p)$ denote the mortality multiplier satisfying,

$$\frac{\mathring{e}^{(\lambda_{ri}(p))}_{x_i}}{e(i)} = 1 + p. \tag{6}$$

Then the relative IDLE, which we denote by $\hat{p}$, is the proportion p which solves (5), with $\lambda_{ri}(p)$ in place of $\lambda_{ai}(d)$. It is the uniform proportional discrepancy of the LEs implied by the discrepancy between the expected term life expectation and its empirical analog.

Since $\lambda_{ai}(\cdot)$ and $\lambda_{ri}(\cdot)$ are defined in terms of the inverse of expected lifetimes as a function of the mortality multiplier, the computation of the IDLE requires the use of a mortality distribution for each insured. Moreover, it is also needed for the calculation of the term lifetimes in (5). In the setting that we work in, the IDLE recommends the use of the reference table used by the LE provider, and the individual mortality multiplier stated in the LE report of each insured.

2.2. Understanding the potential for bias

In the following we will assume that the lifetimes of the insureds follow that resulting from a baseline mortality table and an individual multiplier. The quantities derived from this baseline mortality table will be denoted with a superscript T, *e.g.* as in q^T_x, and λ^T the associated multiplier. For the purpose of isolating the source of bias, let us consider a large policy block consisting of insureds with identically distributed future lifetimes and identically underwritten at the same time, with the same age x, and with the same multiplier. For such a block, an approximation to the DTLE, by using a suitable law of large numbers, is given by

$$\mathring{e}^{T,(\lambda^T)}_{x:\overline{t}|} - \mathring{e}^{(\lambda)}_{x:\overline{t}|}. \tag{7}$$

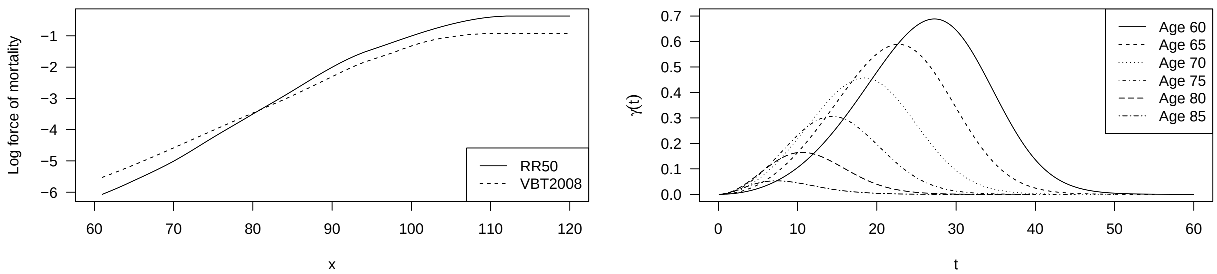
However, the multiplier λ^* derived by equating (7) to zero depends on t whenever the reference table differs in shape from that of the true. In particular, this implies that even for an LE provider with accurate underwriting, the IDLE may conclude with a bias approximately equal to

$$\mathring{e}^{T,(\lambda^T)}_{x} - \mathring{e}^{(\lambda^*)}_{x}. \tag{8}$$

By way of illustration, we will demonstrate this phenomenon in the following setting. We consider male nonsmoker insureds, and for simplicity employ only the ultimate part of the mentioned mortality tables. Suppose that the reference table is the 2008 Valuation Basic Table (VBT) of the Society of Actuaries (SOA) (Society of Actuaries, 2009), whereas the true mortality follows the RR50 sub-table of the VBT 2015 (American Academy of Actuaries and Society of Actuaries, 2018). For brevity, we will refer to these as the VBT08 and the RR50 table, respectively. That these tables differ in shape is seen in Figure 2a. Note that, to ensure parity, we scaled the VBT08 such that its $\mathring{e}_{60}$ matches with that of the RR50.

The difference in Figure 2a can be seen to arise for two reasons. First, relative to its base table, the RR50 table has substantially lower mortality rates at earlier ages, and these rates gradually merge back into that of its base table by age 95. The second is the difference between the 2008 and 2015 versions of the VBT.

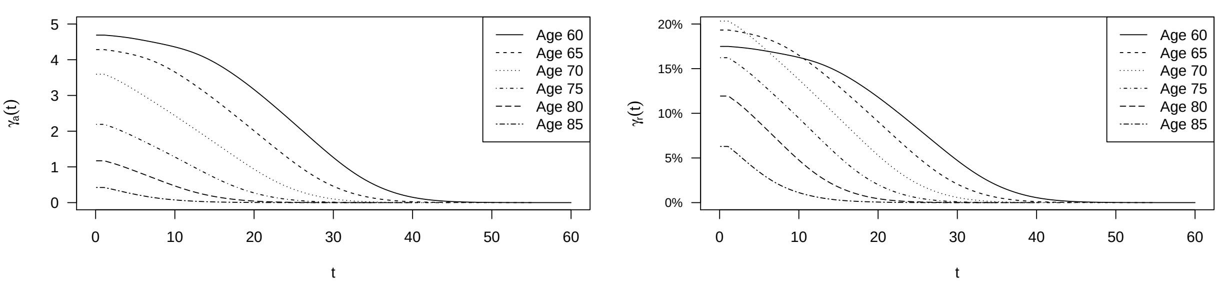
We first demonstrate an alternate view of this phenomenon: that is, even when the two tables are scaled such that their complete life expectancies are equal at a certain age, the resultant term complete life expectancies may differ. Figure 2b shows how the latter difference depends on the term t, for various ages x. We can see that these differences peak at moderate values of t, across ages, and the difference decays with increasing



(a) (b)

Figure 2: (a): Plot of $\log \mu_x^{RR50}$ and $\log \mu_x^{VBT08,(\lambda^*)}$, with λ^* solving $\mathring{e}_{60}^{VBT08,(\lambda^*)} = \mathring{e}_{60}^{RR50}$. (b): Plot of $\gamma(t) := \mathring{e}_{x:\overline{t}|}^{RR50} - \mathring{e}_{x:\overline{t}|}^{VBT08,(\lambda)}$ for various x, with λ solving $\mathring{e}_x^{VBT08,(\lambda)} = \mathring{e}_x^{RR50}$.

age across all t. This renders insight into the distribution of the underwriting age and the age of the policy of a portfolio block that are likely to bias the IDLE. Interestingly, Figure 3a, which is analogous to Figure 2b but where we instead match the term complete life expectancies, exhibits a very different pattern. This represents the approximate bias in the IDLE if everyone in the sample entered at the same age, and were observed for the same amount of time. Figure 3b is the relative version of Figure 3a. We can see that the largest discrepancies are given by the smallest ages, and for the smallest times since entry. In general, due to differing entry ages and times since entry, the bias in the IDLE can be seen as a complex average of these discrepancies.



(a) (b)

Figure 3: (a) Plot of $\gamma_a(t) := \mathring{e}_x^{VBT08,(\lambda_t^*)} - \mathring{e}_x^{RR50}$; (b) Plot of $\gamma_r(t) := \{\mathring{e}_x^{VBT08,(\lambda_t^*)} - \mathring{e}_x^{RR50}\}/\mathring{e}_x^{RR50}$, with λ_t^* solving $\mathring{e}_{x:\overline{t}|}^{VBT08,(\lambda_t^*)} = \mathring{e}_{x:\overline{t}|}^{RR50}$.

2.3. Simulation study

Below we describe the setup of our simulation study which is designed around a realistic life settlements portfolio of size 5000 that is less seasoned and has a relatively younger underwriting age distribution. Our choice of the distribution of the underwriting ages is motivated by Januário and Naik (2016); in particular, by their Figure 4 (d). In their study, they consider approximately 9000 policies funded by Coventry First® originated from across the 50 states, and dated between 2001 and 2011. Their distribution is well-approximated by a truncated normal on the range of 60-90; we consider the same shape but, guided by the earlier discussion, instead on the range 60-80. To be precise, we simulate the age from $TN(\mu = 70, \sigma^2 = 5^2, (a, b) = (60, 80))$, where $TN(\mu, \sigma^2, (a, b))$ denotes the normal distribution with mean μ and variance σ^2 truncated to the interval (a, b).

Since there is a lack of description in the literature of the distribution of mortality multipliers prevalent in the market, we decided to use the non-informative uniform distribution for λ^T. In Table 3 of Braun and Xu (2017), summary statistics of life expectancy assessments of a recent portfolio is given. Since the portfolio studied in Braun et al. (2015) is similarly recent, we backed out a uniform distribution for the mortality multipliers which combined with the age distribution in Braun et al. (2015) results in similar summary statistics as in Braun and Xu (2017). The so chosen distribution is a uniform distribution on the interval $(1.5, 4.5)$.

We assume that the portfolio is 15 years old. This is reasonable as the life settlements market only started to take off in the late 90's when the viatical settlements markets became unprofitable. We also assume that the times at which insureds are underwritten are distributed uniformly across the life of the policy block. In other words, we assume that underwriting time follows $Unif(0, 15)$.

Once the underwriting ages and mortality multipliers are simulated, the death times are then generated independently from the suitably scaled RR50 table. We summarize our simulation setup in Simulation Setup 1.

Simulation Setup 1: Studying the bias of the IDLE

i. Number of insureds in the portfolio: $n = 5000$.

ii. Underwriting ages: $x_i \overset{\text{i.i.d.}}{\sim} TN(\mu = 70, \sigma^2 = 5^2, (a, b) = (60, 80))$.

iii. Mortality multipliers: $\lambda_i^T \overset{\text{i.i.d.}}{\sim} Unif(1.5, 4.5)$.

iv. Times of death: Independently from λ_i^T scaled RR50 table.

v. Censoring times: $t_i \overset{\text{i.i.d.}}{\sim} Unif(0, 15)$.

vi. IDLE computation:

   (a) Reference table: VBT08.

   (b) Life expectancy assessments: $e(i)$ from the λ_i^T-scaled RR50 .

We simulate 1000 such portfolios, and for each of these calculate both the absolute and relative IDLEs using the VBT08 as the reference table. The simulation and computation is done in the `R` statistical software package (R Core Team, 2019).

The results for both the absolute and relative IDLE are as shown in Figure 4. As the DTLE is asymptotically normal, standard theory for Z-estimators (see e.g. van der Vaart (1998), Ch. 5) guarantees that the IDLE too is asymptotically normal. Figure 4 shows that this is indeed the case for both versions of the metric.

In the following, we let $e^T(i)$ denote the complete life expectancy resulting from the true table being scaled by λ_i^T. Given that $e(i) = e^T(i)$ for each i, if the IDLE is indeed unbiased then the simulated densities should be centered around zero. However, Figure 4 shows that this is not the case for both the absolute and relative versions of the IDLE. Taken at face value, the IDLE suggests that the LE provider's assessments are on average long by 0.74 years for the absolute version, or by 6.5% for the relative version. It is noteworthy that the two versions are consistently biased, since the mean life expectancy in our simulation setup is 12 years. For interpreting these biases, we note that in Bauer et al. (2018) an absolute IDLE of 1 year was deemed as corresponding to a more aggressive LE provider. The IDLE metric hence erroneously suggests that the neutral LE provider of our simulation setup is moderately aggressive.

While the pair of tables considered above rendered the IDLE positively biased, we note that this need not always be the case. Figure 2a shows that the ratio of the force of mortality $\mu_x^{RR50,(\lambda^*)}/\mu_x^{VBT08}$ is

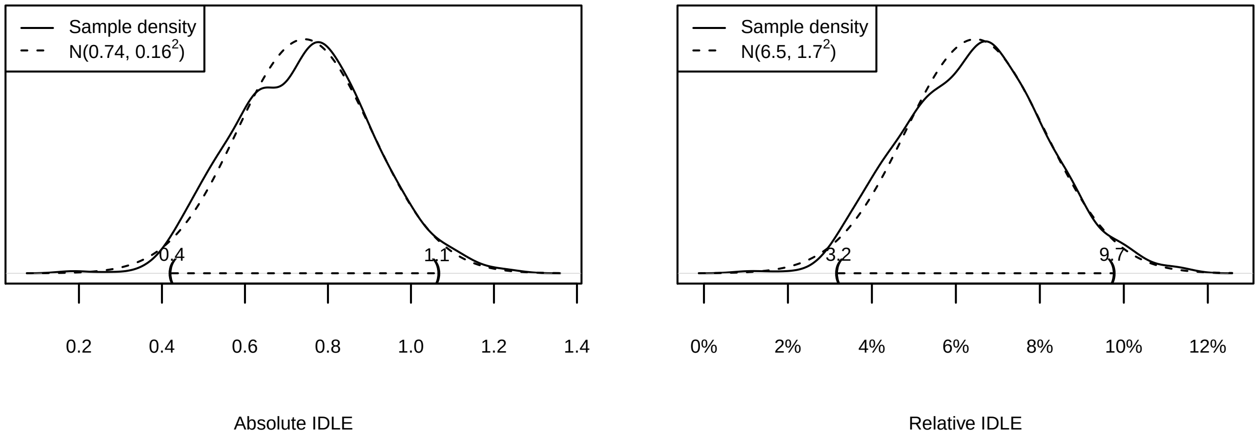
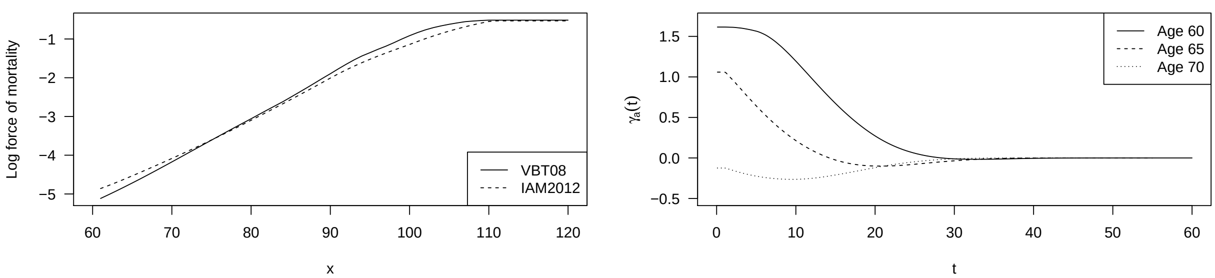


Figure 4: Density plots of the absolute and relative IDLE together with the fitted normal distribution. The interval represents the middle 95% of the fitted distribution.

monotonically increasing with age. When we scale these tables such that they match in their term complete life expectancies $\mathring{e}_{x:\overline{t}|}$, we are ignoring mortality for ages larger than $x+t$. Hence when the ratio of the forces of mortality is monotonically increasing, the higher force of mortality of the RR50 for larger ages leads to smaller survival probabilities ${}_sp_x$, and hence

$$e_x^{VBT08,(\lambda^*)} - e_x^{RR50} = \int_t^\infty ({}_sp_x^{VBT08,(\lambda^*)} - {}_sp_x^{RR50})ds > 0. \tag{9}$$

When the hazard ratio is not monotone, as in the case of VBT08 vs the 2012 Individual Annuity Reserving Table[1] (henceforth abbreviated as IAM12), Figure 5b suggests that the bias in the IDLE can be both positive or negative, depending on the distribution of the ages and the censoring times in the block.



(a) (b)

Figure 5: (a): Comparison of μ_x^{VBT08} with $\mu_x^{IAM12,(\lambda^*)}$, where λ^* satisfies $\mathring{e}_{60}^{IAM12,(\lambda^*)} = \mathring{e}_{60}^{VBT08}$. (b): Plot of $\mathring{e}_x^{IAM12,(\lambda_t^*)} - \mathring{e}_x^{VBT08}$, where λ_t^* satisfies $\mathring{e}_{x:\overline{t}|}^{IAM12,(\lambda_t^*)} = \mathring{e}_{x:\overline{t}|}^{VBT08}$.

In summary the potential for bias makes it difficult to interpret the IDLE, while also affecting the validity of the associated confidence intervals and hypothesis tests. This motivates our search for a metric which is more robust to the choice of the reference table.

[1] While the IAM12 is unlikely to be used in the life settlements market, we use it for the limited use of lending insight into the direction of the bias.

3. Towards a more Robust Metric for LE Assessments

We begin by presenting an alternate viewpoint of the IDLE that helps in naturally arriving at the framework of our metric. The central idea behind the IDLE is to attribute the observed DTLE (see (3)) to misspecifications in the mortality multipliers as the underlying baseline mortality is taken to be proportional to that of the reference table. Due to the inherent randomness of observed lifetimes, and unidimensionality of the DTLE, these multipliers are *corrected* to account for the non-zero observed DTLE with the restriction that they lie on a uniparameteric curve. Finally, the DLE is estimated by the average difference in the complete life expectancies derived using the *corrected* and the LE provider's multipliers. Note that the estimated DLE can be reported directly or as a proportion relative to the average life expectancy assessments. There are two versions of the IDLE corresponding to two ways of parametrizing the corrected multipliers - the absolute (as in (4)) and the relative (as in (6)) versions[2].

As noted in the previous section, the IDLE may be biased in settings where the underlying baseline mortality is not proportional to that of the reference table. So instead of using the data to solely infer (or *correct*) the mortality multipliers (corresponding to the reference table), we instead use the data to infer both the mortality multipliers and the underlying baseline mortality. We parametrically estimate the mortality multipliers, and nonparametrically estimate the baseline mortality. The reference table essentially only appears in the parametric model for the mortality multipliers. In a sense, we reduce the assumptions we make in designing our metric by adopting a more data driven approach.

A standing assumption we make is that for some $\lambda_i^T > 0$,

$$\mu_i^T(\cdot) = \lambda_i^T \cdot \mu_0^T(\cdot),$$

where $\mu_0^T(\cdot)$ is the baseline force of mortality, and $\mu_i^T(\cdot)$ is that corresponding to the i-th insured. In the discussion section we point to a way to relax even this assumption.

3.1. *A Semiparametric Model for* (λ^T, μ_0^T)

Since the group of insureds is heterogeneous, our plan is to leverage the information contained in the LE provider's mortality multiplier (λ) to construct a parsimonious parametric model for the true multiplier (λ^T). Towards specifying a model, we consider an LE provider whose life expectancy estimates are accurate. In this setting, the relationship between λ and λ^T is a function of x, the reference baseline table and the true baseline table. We first consider the simplistic case of $\mu_0(\cdot) \propto \mu_0^T(\cdot)$ which results in $\lambda \propto \lambda^T$. In this case, the model

$$\log(\lambda^T) = \beta \log(\lambda),$$

with $\beta > 0$ would serve well. This is so as $\beta = 1$ would be a perfect fit when μ_0^T is nonparametrically estimated. Outside of this simplistic case, the relationship between λ^T and λ is very much dependent on the reference and true mortality tables. Based on our experience with different pairs of tables, $\log \lambda$ is approximately linear in $\log \lambda^T$ for fixed x, *i.e.*

$$\log \lambda^T \approx \theta_0(x) + \theta_1(x) \log \lambda, \tag{10}$$

for some functions $\theta_0(\cdot)$ and $\theta_1(\cdot)$. In the case when the VBT08 is the reference table and the RR50 is the true table, Figure 6 demonstrates this phenomenon. In order to investigate the intercept and slope functions $\theta_0(\cdot)$ and $\theta_1(\cdot)$, for ages between 60 and 90, we obtain a least squares of $\log(\lambda^T)$ on $\log \lambda$ for each such age. Plots of these functions are given in Figure 7. We note that both θ_0 and θ_1 appear to be as well approximated by a straight line. The slope at age 60 of about 1.18 is substantially higher than that at age 90 which is close

[2]We disconnect the manner in which the estimated DLE is reported, absolute or relative, and the type of uniparametrization enforced on the corrected multipliers, absolute or relative. This is unlike in Bauer et al. (2018) where the reporting is tied to the parametrization.

to 1; this suggests that we may get a better fit by allowing θ_1 to vary with age. The preceding discussion suggests the following model for $\log \lambda^T$:

$$\log \lambda^T = \underbrace{\beta_{00} + \beta_{01} x}_{\theta_0(x)} + \underbrace{\beta_{10} \log \lambda + \beta_{11} x \cdot \log \lambda}_{\theta_1(x) \cdot \log \lambda}.$$

Given that we are estimating the baseline μ_0^T nonparametrically, the constant term β_{00} is redundant. This results in concluding that a plausible parametric model for λ^T is given by

$$\log \lambda^T = \beta_{01} x + \beta_{10} \log \lambda + \beta_{11} x \cdot \log \lambda. \tag{11}$$

The above suggests the Cox proportional hazards model for (λ^T, μ_0^T).

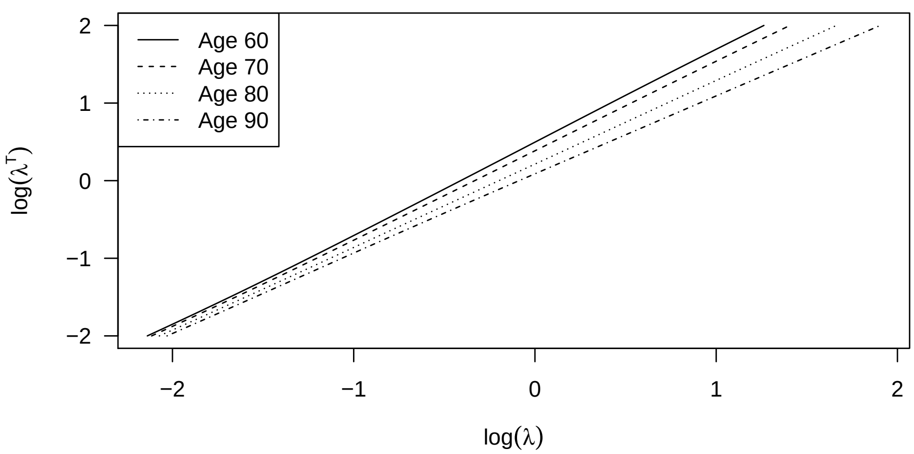


Figure 6: Plot of $\log \lambda^T$ with λ^T defined by $\mathring{e}_x^{RR50,(\lambda^T)} = \mathring{e}_x^{VBT08,(\lambda)}$.

Under the Cox proportional hazards model each person's force of mortality is given by

$$\mu(x) = \exp(\mathbf{z}'\boldsymbol{\beta})\mu_0(x) \tag{12}$$

where $\mathbf{z}$ is the $p \times 1$ vector of covariates, and $\boldsymbol{\beta}$ is an unknown vector of coefficients that is to be estimated. In other words, this assumes that the mortality multiplier λ is log-linear in the covariate vector $\mathbf{z}$. Note that (11) suggests the use of $(x, \log \lambda, x \log \lambda)$ as covariates. Statistical inference begins by deriving an estimate of $\boldsymbol{\beta}$ by maximizing its partial likelihood, and then the baseline mortality is estimated. The inference procedure easily accommodates left-truncation (differing underwriting ages) and right-censoring (finite portfolio age). We refer to standard references such Therneau and Grambsch (2000) and Kleinbaum and Mitchel (2012) for details on the estimation process.

Given the estimates of the baseline force of mortality and the mortality multipliers, the estimate of the life expectancy for the i-th insured, which we denote $\hat{e}(i)$, can then be obtained. This allows for the direct estimation of the DLE, by calculating their average discrepancy from the life expectancy assessments, which is our proposed metric: the *Estimated Difference in Life Expectancies* (EDLE). Specifically, our proposed EDLE metric is given by,

$$\frac{1}{n} \sum_{i=1}^{n} (\hat{e}(i) - e(i)). \tag{13}$$

3.2. **EDLE *in Practice***

While one of the strengths of the EDLE is its natural formulation, practical implementation requires a further tweak for reasons as we now explain. A practical issue in estimating survival models arises from the lack of

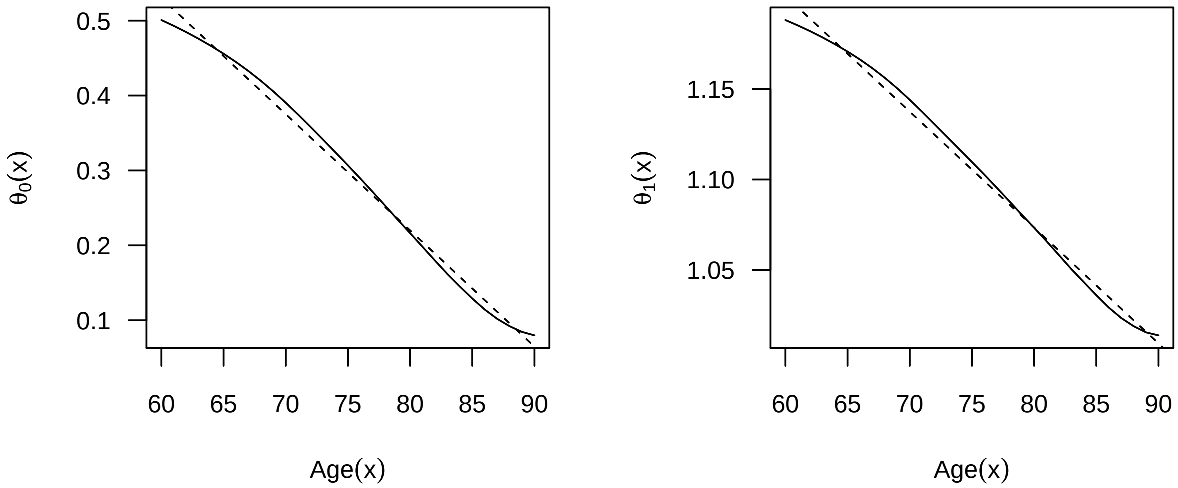


Figure 7: Plots of $\theta_0(\cdot)$ and $\theta_1(\cdot)$ along with their straight line fits.

mortality data at advanced ages; in our case it is either due to the short life expectancies of the insureds in the life settlements markets, or due to censoring. To accommodate for this we use the estimated mortality up to some age x^*, for which we deem the estimate remains credible. Then we resort to the life expectancy estimate for the i-th insured given by

$$\hat{e}(i) = \hat{e}_{\overline{(x^*-x_i)}|}(i) + \hat{S}_{(i)}(x^*)\check{e}_{x^*}(i), \tag{14}$$

where $\hat{e}_{\overline{t}|}(i)$, $\hat{S}_{(i)}(\cdot)$ denote the estimated term life expectancy and the estimated survival function for the i-th insured, respectively. Also, $\check{e}_{x^*}(i)$ is an estimate for $e_{x^*}(i)$ derived using an alternate methodology that is less data and more assumption driven. Some options for specifying the latter include to simply using $\mathring{e}_{x^*}^{(\lambda_i)}$, or using either $\mathring{e}_{x^*}^{(\lambda_{ri}(d))}$ or $\mathring{e}_{x^*}^{(\lambda_{ri}(p))}$ arising from the IDLE methodology.

The use of a less data driven $\check{e}_{x^*}(i)$ as an estimate of the $\mathring{e}_{x^*}^{T(\lambda_i^T)}$ potentially can bias $\hat{e}(i)$. The extent of this bias depends on the average contribution of the residual life beyond x^* to the whole life expectancies of the insureds. When the mortality data contains information up to a sufficiently large age (95 or above), this is not an issue as is confirmed by our later described simulation results. This can fail to happen for a policy block with predominantly recently underwritten policies over a narrow range of underwritten ages. But in such cases, Figures 1 and 5 in Bauer et al. (2018) shows that even the A/E ratio and the IDLE tend to be less precise. Finally, from our experiments we note that the performance of $\hat{e}(i)$ is not materially susceptible to the choice of a reasonable sample driven methodology for determining x^*.

4. Asymptotic Results for the EDLE

In this section, we establish asymptotic results for the EDLE that enable construction of confidence intervals for the EDLE. These can be used to make judgements on the quality of the life expectancy assessments from a LE provider. Classical results for the Cox model guarantee the joint asymptotic normality of the estimates for $\boldsymbol{\beta}$ and the integrated force of mortality for a wide range of Cox models (Andersen and Gill, 1982; Andersen et al., 1992). Since the life expectancies are a functional of these two quantities, its asymptotic normality follows from employing a suitable functional delta method. This yields asymptotic normality of the EDLE as well. We also give a natural consistent estimator for the asymptotic variance of the EDLE. We give a precise statement of these results in this section with their proofs relegated to the Appendix.

Notation for the Asymptotic Results: In our setting, P^0 denotes the joint probability measure of $(x, \mathbf{z}, T_0, t)$ where T_0 denotes the age at death. We also assume that $x \sim G$, $\mathbf{z}|x \sim K_x$, $\mu_{T_0|\mathbf{z}} = \exp(\mathbf{z}' \boldsymbol{\beta}_0) \cdot \mu_0$, and $t \sim H$, where the K_x's for each $x > 0$, as well as G and H, are each cumulative distribution functions. Also, we assume that conditional on $\mathbf{z}$, we have mutual independence of x, t and T_0; in other words, truncation and censoring are nonselective.

Now let $P := P^0 | T_0 > x$, and suppose we observe an i.i.d. random sample of $(x, \mathbf{z}, T_x \wedge t, \delta)$ from P, where $\delta = \mathbf{1}(T_x < t)$. Further, we restrict our attention to some age interval $\mathcal{X} = [x_l, x_u]$, with $x_l > 0$, $x_u < \infty$. Let $\hat{\boldsymbol{\beta}}$, $\hat{\mathrm{M}}_0(\cdot)$ denote the estimates of the coefficients $\boldsymbol{\beta}$ and the integrated force of mortality $\mathrm{M}_0^T(y) := \int_0^y \mu_0^T(s)ds$ from the Cox proportional hazards model, respectively. For $x \in \mathbb{R}$, let $x_+ := \max(x, 0)$, and for a vector $\mathbf{x}$ we let

$$\mathbf{x}^{\otimes(0)} = 1; \qquad \mathbf{x}^{\otimes(1)} = \mathbf{x}; \qquad \mathbf{x}^{\otimes(2)} = \mathbf{x}\mathbf{x}'.$$

We define the following quantities that will be needed later:

$$\lambda^T = \exp(\mathbf{z}' \boldsymbol{\beta}_0); \qquad Y(y) = \mathbf{1}(x < y, x + T_x \wedge t > y);$$

$$S_x(y; \mathbf{z}) = \exp\left\{-\lambda^T (\mathrm{M}_0(y) - \mathrm{M}_0(x))_+\right\}; \qquad \mathbf{s}^{(m)}(y, \boldsymbol{\beta}_0) = \mathbb{E}[\mathbf{z}^{\otimes(m)} \lambda^T Y(y)], \quad m = 0, 1, 2;$$

$$\mathbf{e}(y, \boldsymbol{\beta}_0) = \frac{\mathbf{s}^{(1)}(y, \boldsymbol{\beta}_0)}{s^{(0)}(y, \boldsymbol{\beta}_0)}; \qquad \mathbf{v}(y, \boldsymbol{\beta}_0) = \frac{\mathbf{s}^{(2)}(y, \boldsymbol{\beta}_0)}{s^{(0)}(y, \boldsymbol{\beta}_0)} - \mathbf{e}^{\otimes(2)}(y, \boldsymbol{\beta}_0);$$

$$\hat{\lambda}_i = \exp(\mathbf{z}_i' \hat{\boldsymbol{\beta}}); \qquad \hat{S}_x(y; \mathbf{z}) = \exp\left\{- \exp(\mathbf{z}' \hat{\boldsymbol{\beta}})(\hat{\mathrm{M}}_0(y) - \hat{\mathrm{M}}_0(x))_+\right\};$$

$$Y_i(y) = \mathbf{1}(x_i < y, x_i + T(i) \wedge t_i > y); \qquad \mathbf{S}^{(0)}(y, \boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{z}_i^{\otimes(m)} \exp(\mathbf{z}_i' \boldsymbol{\beta}) Y_i(y), \quad m = 0, 1, 2;$$

$$\mathbf{E}(y, \boldsymbol{\beta}) = \frac{\mathbf{S}^{(1)}(y, \boldsymbol{\beta})}{S^{(0)}(y, \boldsymbol{\beta})}; \qquad \mathbf{V}(y, \boldsymbol{\beta}) = \frac{\mathbf{S}^{(2)}(y, \boldsymbol{\beta})}{S^{(0)}(y, \boldsymbol{\beta})} - \mathbf{E}^{\otimes(2)}(y, \boldsymbol{\beta})$$

$$N(y) = \sum_{i=1}^n \delta_i \mathbf{1}(x_i + T(i) > y)$$

Note that $S^{(0)}(y, \boldsymbol{\beta})$ represents the multiple of the hazard function in effect at time y, and

$$n^{-1}\, \mathbb{E}[S^{(0)}(y, \boldsymbol{\beta})] = s^{(0)}(y, \boldsymbol{\beta}).$$

In a sense, $s^{(0)}$ represents the amount of data available for the estimation of mortality rates at age y, with the estimation uncertainty for the cumulative hazard M_0 being tied to the reciprocal of this quantity.

Assumptions:

(A1) $\int_{\mathcal{X}} \mu_0(x) dx < \infty$.

(A2) There exists a neighborhood B of $\boldsymbol{\beta}_0$ such that

$$\mathbb{E}\left\{\sup_{y \in \mathcal{X}, \boldsymbol{\beta} \in B} Y(y) \|\mathbf{z}\|^2 \exp(\mathbf{z}^T \boldsymbol{\beta})\right\} < \infty.$$

(A3) $Pr(x + t > x_u) > 0$.

(A4) $\mathbb{E}\left[\|\mathbf{z}\|^m \exp(\mathbf{z}^T \boldsymbol{\beta}) | x = x_l\right] \in (0, \infty)$ for $m = 0, 1, 2$.

(A5) $\tau^2 = \int_{\mathcal{X}} \left(\mathbb{E}_{x,\mathbf{z}} \left\{\lambda^T S_x(y; \mathbf{z}) \mathring{e}_y^{T,(\lambda^T)} \mathbf{1}(x < y)\right\}\right)^2 (\mu_0(y)/s^{(0)}(y, \boldsymbol{\beta}_0)) dy < \infty$.

(A6) The expectations $\mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}]$ and $\mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}\mathbf{z}]$ exist and are finite.

A nonzero probability of a person surviving the age interval $\mathcal{X}$ guarantees (A1), and hence is satisfied in our setting by choosing $\mathcal{X}$ suitably. Assumptions (A2), (A3) and (A4) are easily interpreted mild conditions that are satisfied in our setting, and these are akin to those in Andersen and Gill (1982).

The asymptotic variance of the EDLE naturally decomposes into two parts; the first arising from the variability of $\hat{\boldsymbol{\beta}}$, and the second arising from the variability inherent in estimating M_0 given the knowledge of $\boldsymbol{\beta}$. The quantity τ^2 of (A5) represents the latter. Under (A1) and (A3), (A5) essentially requires that

$$\lim_{y \to x_l} \frac{\left(\mathbb{E}_{x,\mathbf{z}}\left\{\lambda^T S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}\mathbf{1}(x<y)\right\}\right)^2}{s^{(0)}(y,\boldsymbol{\beta}_0)} < \infty \tag{15}$$

Finally, (A6) is another mild condition that is required for the consistency of the sample estimator of the asymptotic variance.

The asymptotic normality of the EDLE is established by the following result.

Theorem 1. *Given that assumptions (A1) - (A6) hold,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{e}(i) - \mathring{e}_{x_i}^{T,(\lambda_i^T)}) \xrightarrow{\mathrm{d}} N(0, \mathbf{k}_{\boldsymbol{\beta}}^T \Sigma_{\mathcal{X}}^{-1}\mathbf{k}_{\boldsymbol{\beta}} + \tau^2) \tag{16}$$

where

$$\mathbf{k}_{\boldsymbol{\beta}} = \int_{\mathcal{X}} \mathbb{E}_{x,\mathbf{z}_0}\left\{\lambda S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}(\mathbf{z} - \mathbf{e}(y,\boldsymbol{\beta}_0))\mathbf{1}(x<y)\right\}\mu_0(y)dy,$$

and

$$\Sigma_{\mathcal{X}} = \int_{\mathcal{X}} \mathbf{v}(y,\boldsymbol{\beta}_0)s^{(0)}(y,\boldsymbol{\beta}_0)dy.$$

Moreover, a consistent estimator of this variance is given by $\hat{\mathbf{k}}_{\boldsymbol{\beta}}^T\hat{\Sigma}_{\mathcal{X}}^{-1}\hat{\mathbf{k}}_{\boldsymbol{\beta}} + \hat{\tau}^2$, *where*

$$\hat{k}_{\boldsymbol{\beta}} = \frac{1}{n}\int_{\mathcal{X}} \frac{\sum_{i:x_i<y}\hat{\lambda}_i\hat{S}_{x_i}(y;\mathbf{z}_i)\hat{e}_y^{(\hat{\lambda}_i)}(\mathbf{z}_i - \mathbf{E}(y,\hat{\boldsymbol{\beta}}))}{S^{(0)}(y,\hat{\boldsymbol{\beta}})}dN(y),$$

$$\hat{\tau}^2 = \frac{1}{n}\int_{\mathcal{X}}\left(\frac{\sum_{i:x_i<y}\hat{\lambda}_i\hat{S}_{x_i}(y;\mathbf{z}_i)\hat{e}_y^{(\hat{\lambda}_i)}}{S^{(0)}(y,\hat{\boldsymbol{\beta}})}\right)^2 dN(y), \quad \text{and} \quad \hat{\Sigma}_{\mathcal{X}} = \frac{1}{n}\int_{\mathcal{X}}\mathbf{V}(y,\hat{\boldsymbol{\beta}})dN(y).$$

An analogous result for the version of the EDLE in (14) is given below. For this result, (A5'), a weakening of assumption (A5), suffices:

(A5') With $\mathcal{X}^* = [0, x^*]$,

$$\tau^2 = \int_{\mathcal{X}^*}\left(\mathbb{E}_{x,\mathbf{z}}\left\{\lambda^T S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}\mathbf{1}(x<y)\right\}\right)^2(\mu_0(y)/s^{(0)}(y,\boldsymbol{\beta}_0))dy < \infty$$

Theorem 2 assumes that the x^* in (14) is nonrandom, but it easily extends to the case where the variability in x^* is $o(n^{-1/2})$. For the conditions of Theorem 2 to be satisfied, the choice of x^* must be such that the probability of having an insured in the policy block surviving to age x^* is positive.

Theorem 2. *Suppose the estimated life expectancies are calculated according to* (14)*, where x^* is fixed. Then if $\breve{e}_{x^*}(i) := \mathring{e}_{x^*}^{T,(\lambda_i^T)}$ for all i, and if (A1) through (A4), (A5') and (A6) hold, then the conclusions of Theorem 1 hold with $\mathcal{X}^*$ in place of $\mathcal{X}$ in the definitions of $\mathbf{k}_{\boldsymbol{\beta}}$, τ^2, $\hat{\mathbf{k}}_{\boldsymbol{\beta}}$ and $\hat{\tau}^2$, and with the definition of $\hat{e}(i)$ given by* (14).

5. Simulation Study of the EDLE

In this section the goal of our simulation study is threefold. First, we use Simulation Setup 1 to study the performance of the EDLE in the case when the reference baseline table differs in shape from the baseline true table. Second, we study the performance of the IDLE and the EDLE in the case when the reference baseline table is proportional to the baseline true table. We do this as the IDLE methodology is expected to do better than the EDLE in this setting, and comparable performance by the EDLE would be supportive of its worthiness as a metric. Finally, we study the coverage levels of the nominal asymptotic confidence intervals and proximity to normality of the finite sample distribution of the EDLE. Since the IDLE is developed in two versions, we introduce a relative version of the EDLE as given by,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{e}(i)-e(i)}{e(i)}\right). \tag{17}$$

We study both the absolute and the relative versions of the EDLE below.

5.1. Performance of the EDLE: Tables of Differing Shapes

We begin by recalling that (11) and the discussion of subsection 3.1 were suggestive of a Cox model with covariates $(x, \log\lambda, x\log\lambda)$. We consider the fit of all of the sub-models towards recommending a simpler model. The simulation setup we employ for this purpose is that of subsection 2.3, namely Simulation Setup 1, which incorporates a reference baseline table that differs in shape from that of the baseline true table and under which the IDLE exhibited material bias. The sub-models we consider, to be specific, correspond to the following choice of covariates, and these models were fit using the `survival` package in `R` (Therneau and Grambsch, 2000):

1. Using only x as a covariate
2. Using only $\log\lambda$ as a covariate
3. Adding x as a covariate to the previous model
4. Further adding the interaction term $x\cdot\log\lambda$

Model fitted			
x only	$\log\lambda$ only	$(\log\lambda,\ x)$	$(\log\lambda,\ x,\ x\cdot\log\lambda)$
29,999	29,808	29,801	29,803

Table 1: AIC values for the four sub-models.

Since we are working with simulated data, we fit these models to a portfolio corresponding to a single simulation iteration. We note that the conclusions drawn below for this single run were invariant across most other runs. In Table 1 we present the AIC values for these models. As one would expect, the model with only x as a covariate is the worst performer, and the model with covariates $(\log\lambda, x)$ beats that which only uses $\log\lambda$ as a covariate. Interestingly, contrary to what one might suspect from Figure 7, the interaction term does not significantly contribute to the model fit. Hence we will focus on the parsimonious model using only $(\log\lambda, x)$ as the covariates. Table 2 pertains to the fit of this model. We see that both the coefficients are significant. Figure 8 shows the associated Schoenfeld residuals for these two covariates, and they visually support the proportional force of mortality assumption. The associated significance tests, see Table 3, that test for these residuals being significantly correlated with the transformed time (see Therneau and Grambsch (1994)) agrees with this assessment as well.

Covariate	Coefficient	Std. Err.	p-value
$\log \lambda$	1.163	0.085	$< 2 \cdot 10^{-16}$
x	-0.019	0.0065	0.0033

Table 2: Model fit for the Cox model with covariates ($\log \lambda$, x).

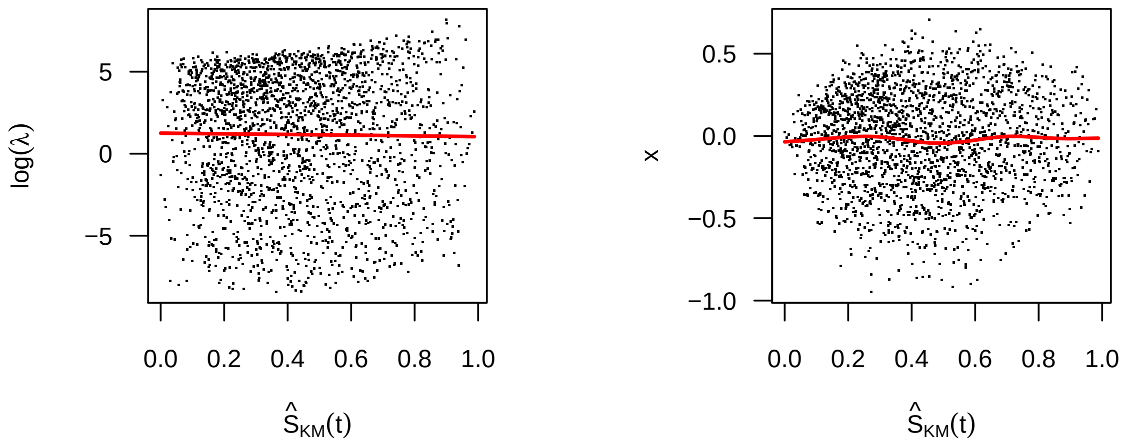


Figure 8: Schoenfeld residuals for the covariates in the ($\log \lambda$, x) model.

5.2. Comparison of IDLE with EDLE

While the results of section 5.1 suggest the $(\log \lambda, x)$ model as one providing a good fit to our data, in the sequel, we will consider the EDLE under both this and the model that uses only $\log \lambda$ as a covariate. This is to demonstrate that the EDLE performance is dependent on choosing a well fitting model.

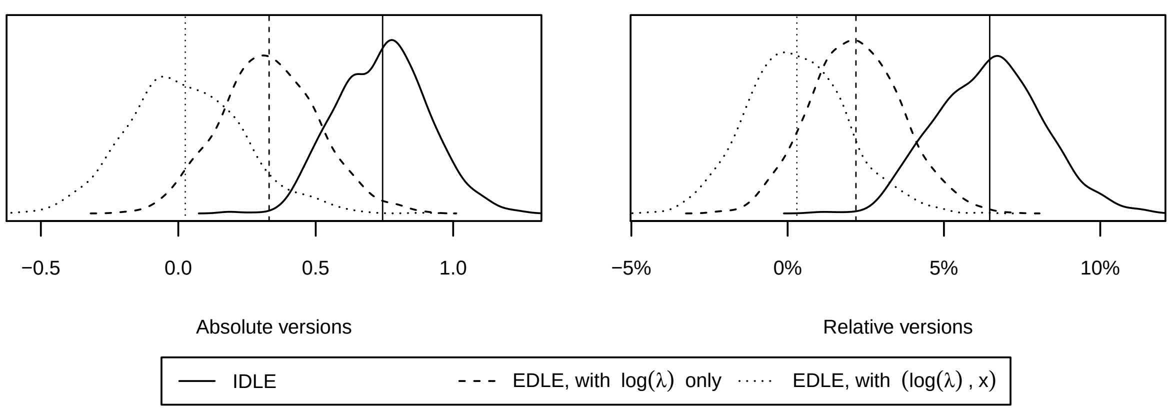


Figure 9: Comparison of the versions of the IDLE and the EDLE - sample densities where the means are marked.

Figure 9 shows the comparison of the absolute and relative IDLE and EDLE under these models. Recall that the IDLE biases were 0.74 years for the absolute version, or 6.5% for the relative version. Using the EDLE corresponding to the model with only $\log \lambda$ as a covariate results in incurring only half this bias for the absolute version, or only one-third of the bias for the relative version. This means that the LE provider's multipliers are indeed informative about the mortality level of each person, albeit unsatisfactorily as the lone covariate - a significant portion of the bias still remains. However, with the inclusion of the age as an additional covariate, the EDLE is almost unbiased across both versions. The models with more covariates appear to be slightly more variable; the extent to which this happens for the relative EDLE is less than that

Covariate	Correlation	Chi-square	p-value
$\log\lambda$	-0.012	0.305	0.581
x	-0.004	0.026	0.820

Table 3: Significance tests for the Schoenfeld residuals.

for the absolute version. Interestingly, while the absolute EDLE appears more variable with the addition of the covariate, the relative EDLE exhibits the opposite effect. Overall, the evidence suggests that the EDLE is indeed more robust to the misspecification of the reference table compared to the IDLE, allowing us to interpret it even under uncertainty about the shape of the underlying baseline mortality. The fitted density to the simulate EDLEs and IDLEs in Figure 9 suggest that the bias of the EDLE employing the $(\log\lambda, x)$ model is negligible compared to that of the IDLE.

We also investigated if the use of splines improved the model fit beyond that afforded by the interaction model. Specifically, we explored if the use of splines to approximate $\theta_0(x)$ of (10) gave a better model fit. However, the EDLE with the splines was virtually indistinguishable from that under the model with only $\log\lambda$ and x, therefore we do not report on such models here.

That the EDLE is more robust towards a misspecified reference table begs an investigation of its performance relative to the IDLE when the reference table and the underlying table are similarly shaped. To address this, we conduct a simulation study using Simulation Setup 2. We again simulate $1,000$ portfolios, and compute the IDLE and EDLE, for both of their respective absolute and relative versions, for each portfolio.

Simulation Setup 2: Comparing the IDLE and EDLE under a correctly specified reference table

Items (i), (iii), (iv) and (vi) are identical to Simulation Setup 1.

ii. Underwriting ages: $x_i \overset{\text{i.i.d.}}{\sim} TN(\mu = 80, \sigma^2 = 5^2, (a, b) = (70, 90))$.

v. Censoring times: $t_i \overset{\text{i.i.d.}}{\sim} Unif(0, 20)$.

vii. IDLE computation:

1. Reference table: RR50.
2. Life expectancy assessments: $\mathring{e}_{x_i}$ from the λ_i^T-scaled RR50 .

viii. EDLE computation: using only $\log\lambda_i$ as a covariate.

Note that in Simulation Setup 2, we have used an age distribution that more closely follows that studied in Braun et al. (2015), reflecting a composition of a more recent portfolio. We have also assumed that the portfolio is more seasoned at 20 years of age, mimicking a portfolio that has existed since the infancy of the life settlements markets. We have changed the setup to better cover the different types of portfolios which potentially exist in the life settlements markets. We note that a similar study using instead Simulation Setup 1 yields similar results.

Figure 10 shows the results for the the absolute and relative versions of the metrics. The biases are negligible, and the spreads look comparable. There is not much loss in the performance that results from using the EDLE over the IDLE - supporting the use of the more robust EDLE.

5.3. Proximity to Normality - Small Portfolio Sizes

We conclude this section by investigating how well our derived asymptotics (specifically, the results of Theorem 1) hold for portfolios with a small to moderate size. For this purpose we consider Simulation Setup 3, a minor modification to Simulation Setup 2. We do this to allow for much longer censoring times such that the whole of the mortality distribution can be theoretically estimated, and hence that the conditions

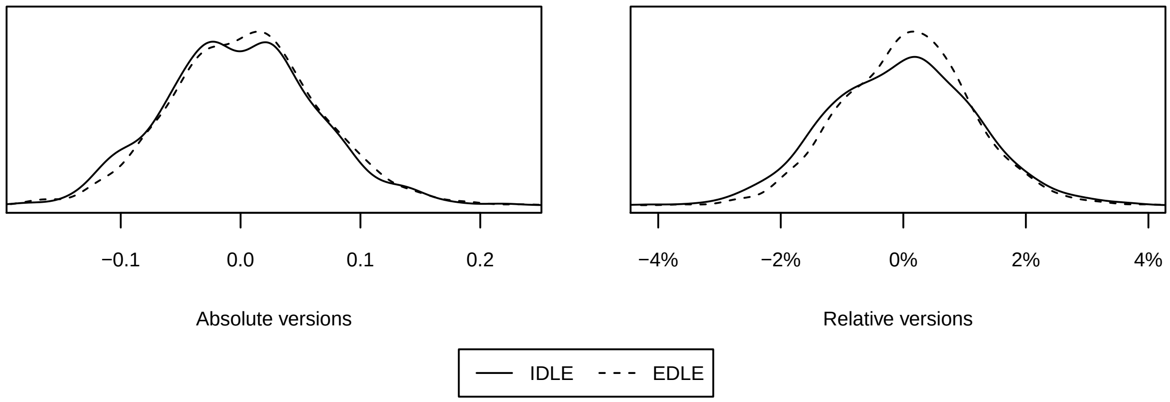


Figure 10: Comparison of the versions of IDLE and EDLE under a correctly specified reference table.

of Theorem 1 hold. For each sample size, we generate $5,000$ such portfolios, and for each portfolio calculate the EDLE and its estimated standard error. The statistic obtained by standardizing the EDLE by this estimate of the standard error is then calculated - we do this simply to facilitate exposition by having a common asymptotic limit. In Figure 11, the sample density of this standardized EDLE is plotted alongside the standard normal density.

Simulation Setup 3: Investigating the asymptotics of the EDLE
Items (ii), (iii), (iv), (vi) and (viii) are identical to Simulation Setup 1. i. Sample size: $n = 50$, $n = 100$,$n = 500$ and $n = 5000$ (four different portfolios). ii. Censoring times: $t_i \sim Unif(0, 40)$.

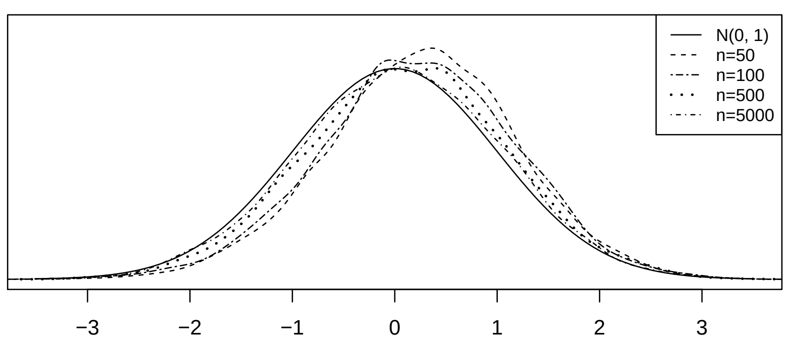


Figure 11: Sample density for the standardized EDLE together with that of $N(0, 1)$.

Figure 11 shows that for the relatively modest portfolio size of $n = 500$ the EDLE is quite close to normal, though it does exhibit a slight positive bias; with $n = 5000$ even this positive bias disappears. This gives us comfort to use confidence intervals and perform hypothesis tests for the EDLE without having to use computationally intensive methods like the bootstrap. Further, that the asymptotic distribution fits so well at small portfolio sizes gives us confidence that the true coverage probabilities/rejection probabilities of these procedures will be close to the nominal level. This is confirmed by the results in Table 4.

Portfolio Size	Nominal Confidence Level		
	90%	95%	99%
50	91.18%	95.70%	99.26%
100	90.90%	95.56%	99.02%
500	90.80%	95.06%	99.14%
5000	90.36%	95.40%	99.32%

Table 4: Simulated coverage probabilities for the nominal asymptotic confidence interval.

6. Summary and Discussion

The IDLE of Bauer et al. (2018) was a significant improvement over the A/E ratio in the sense of being able to maintain its interpretability as the block ages while also retaining the computational simplicity of the A/E ratio. In this paper we argued the need for a metric more robust than the IDLE especially when there is much uncertainty about the shape of the underlying mortality table. Moreover, we suggest a natural metric, the EDLE, arising from fitting a semiparametric survival model to the mortality data. The EDLE we show is robust, a property it derives from the semiparametric nature of the model underlying its construction. And importantly, it maintains its interpretability akin to the IDLE as the block ages, and leverages upon the wide availability of statistical software for fitting the Cox proportional hazards model to retain the simplicity of use that both the A/E ratio and the IDLE possess.

In this article we restricted our attention to metrics which are product neutral, and we were guided by the current standard practice and that such metrics have a wider applicability. It is both natural and reasonable for metrics in this class to be based on estimating the DLE, such as the IDLE and the EDLE. It is worth mention though that as a byproduct of our semiparametric modeling we produce an estimate of the lifetime distribution for each insured. This importantly permits construction of portfolio/product specific metrics which may be easier to interpret for an investor. Also, one can consider product neutral metrics that are more sensitive to deviations between the shape of the LE provider's mortality distribution and what is suggested by the data, This could be done in a natural manner by considering the average distance between the LE provider's mortality distributions and the predicted distributions from the model, as measured through any of the standard metrics on the space of probability measures.

While we consider a single risk class, multiple risk classes, *e.g.* defined by gender and smoking status, can easily be accounted for by using stratified Cox models. The case of latent risk classes is slightly more delicate, but we note that there are generalizations of the Cox model to handle such cases, *e.g.* see Rosen and Tanner (1999). It is of interest to investigate under what conditions that our proposed metric will work well in conjunction with such models.

Unlike Bauer et al. (2018), in this paper we considered only simulation data as we did not find any industry data available in public. This in turn restricted us from exploring a wider choice of covariates, constraining us to using only the mortality multiplier as a summary of such potentially more informative set of covariates. But our semiparametric framework would permit consideration of, for example, significant health conditions that the format of one of the three prominent US LE provider's report includes. As a future work we hope to be able to report on the performance of our proposed methodology on real data incorporating a richer set of covariates.

Acknowledgements

Our interest in this work arose during a visit of Prof. Daniel Bauer to the University of Iowa, and we thank him for useful discussions. The first author is thankful to his audience for the feedback he received from presenting this material at the 23rd Insurance: Mathematics and Economics conference in Munich, Germany, and the 54th Actuarial Research Conference in Indianapolis, US. In particular, he acknowledges useful discussions with Prof. Jochen Russ. The second author acknowledges with gratitude the support from a Society of Actuaries' Center of Actuarial Excellence Research Grant.

References

A. M. Best (2016). Life settlement securitization. http://www3.ambest.com/ambv/ratingmethodology/OpenPDF.aspx?rc=197705.

Actuarial Standards Board (2013). Actuarial standard of practice no. 48: Life settlements mortality. https://www.actuarialstandardsboard.org/wp-content/uploads/2014/02/asop048_1751.pdf.

American Academy of Actuaries and Society of Actuaries (2018, 9). 2015 valuation basic table report. https://www.soa.org/globalassets/assets/files/resources/experience-studies/2018/2015-vbt-report.pdf.

Andersen, P. K., Ø. Borgan, R. D. Gill, and N. Keiding (1992). *Statistical Models Based on Counting Processes*. Springer-Verlag.

Andersen, P. K. and R. D. Gill (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics 10*(4), 1100–1120.

Bauer, D., M. V. Fasano, J. Russ, and N. Zhu (2018). Evaluating life expectancy evaluations. *North American Actuarial Journal 22*(2), 198–209.

Braun, A., S. Affolter, and H. Schmeiser (2015). Life settlement funds: current valuation practices and areas for improvement. *Risk Management and Insurance Review 19*(2), 173–195.

Braun, A., N. Gatzert, and H. Schmeiser (2012). Performance and risks of open-end life settlement funds. *Risk Management and Insurance Review 79*(1), 193–229.

Braun, A. and J. Xu (2017). Dating death: an empirical comparison of medical underwriters in the US life settlements market. In *Asia-Pacific Risk and Insurance Association (APRIA) Annual Conference*.

Davó, C. M. Nuria Bajo, Resco, and M. M. Barroso (2013). Portfolio diversification with life settlements: an empirical analysis applied to mutual funds. *The Geneva Papers 38*, 22–42.

Deloitte (2005). The life settlements market: An actuarial perspective on consumer economic value, research report with the university of connecticut.

Doherty, N. A. and H. J. Singer (2003). The benefits of a secondary market for life insurance policies. *Real Property, Probate and Trust Journal 38*(3), 449–478.

European Life Settlement Association (2013). Life expectancy comparison update. https://www.elsa-sls.org/wp-content/uploads/pdf/life-expectancy-update.pdf.

Gibson, C. (2019, January). Why life settlements will continue to see growth. https://insurancenewsnet.com/innarticle/why-life-settlements-will-continue-to-see-growth.

Januário, A. V. and N. Y. Naik (2016). Testing for adverse selection in life settlements: the secondary market for life insurance policies. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2278299.

Keiding, N. and R. D. Gill (1990). Random truncation models and Markov processes. *The Annals of Statistics 18*(2), 582–602.

Kleinbaum, D. G. and K. Mitchel (2012). *Survival Analysis: A Self-Learning Text* (3 ed.). New York: Springer.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rao, R. R. (1963). *The law of large numbers for* $D[0,1]$*-valued random variables*, Volume 8. Theory of Probability and its Applications.

Rosen, O. and M. Tanner (1999). Mixtures of proportional hazards regression models. *Statistics in Medicine 18*, 1119–1131.

Society of Actuaries (2009, 6). 2008 valuation basic table report. https://www.soa.org/globalassets/assets/files/research/exp-study/research-2008-vbt-report.pdf.

Stone, C. A. and A. Zissu (2006). Securitization of senior life settlements: Managing extension risk. *Journal of Derivatives 13*(3), 66–73.

The Florida Legislature (2019). Life expectancy providers; registration required; denial, suspension. http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0600-0699/0626/Sections/0626.99175.html.

Therneau, T. M. and P. M. Grambsch (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika 81*(3), 515–526.

Therneau, T. M. and P. M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

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

A. Derivation of Asymptotics

For the proofs of Theorems 1 and 2, we crucially rely on classical results in Andersen et al. (1992). Joint asymptotic normality of $\hat{\boldsymbol{\beta}}$ and $\hat{M}_0(\cdot)$ is given by their Theorems VII 2.1, VII 2.2 and VII 2.3, provided their Condition VII 2.1 and VII 2.2 hold, as well as $\int_{\mathcal{X}} \mu_0(y)dy < \infty$. Since the last condition is assumed in our theorem, we need only check their Condition VII 2.1 for our setting. In the following restatement of their conditions, (a) - (e) correspond to their Condition VII 2.1, whereas (f) corresponds to Condition VII 2.2. Finally, (c') is our relaxation of their condition (c) which suffices for the more restricted form of asymptotic normality needed for the proof.

Condition 1. *There exist a neighborhood B of $\boldsymbol{\beta}_0$ and scalar, p-vector and $p \times p$ matrix-valued functions $s^{(0)}$, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$ respectively, defined on $\mathcal{X} \times B$, such that:*

(a)

$$\sup_{\beta \in B, y \in \mathcal{X}} \left\| \frac{1}{n}\mathbf{S}^{(m)}(y,\boldsymbol{\beta}) - \mathbf{s}^{(m)}(y,\boldsymbol{\beta}) \right\| \xrightarrow{\text{p}} 0 \tag{18}$$

(b) *$\mathbf{s}^{(m)}$ is continuous uniformly in $y \in \mathcal{X}$ and bounded on $\mathcal{X} \times B$.*

(c) *$s^{(0)}(\cdot, \boldsymbol{\beta}_0)$ is bounded away from 0.*

(c') *For fixed $\boldsymbol{\beta} \in B$, $s^{(0)}(\cdot,\boldsymbol{\beta}) > 0$ a.e. with respect to the measure with Lebesgue density μ_0, restricted to $\mathcal{X}$. Further, for those $(y,\boldsymbol{\beta})$ such that $s^{(0)}(y,\boldsymbol{\beta}) = 0$, we have*

$$0 < \lim_{x \to y} \frac{s^{(0)}(x,\boldsymbol{\beta})}{s^{(0)}(x,\boldsymbol{\beta}_0)} < \infty \tag{19}$$

$$\lim_{x \to y} \frac{\|\mathbf{s}^{(i)}(x,\boldsymbol{\beta})\|}{s^{(0)}(x,\boldsymbol{\beta})} < \infty \text{ for } i = 1,2 \tag{20}$$

(d) *$\mathbf{s}^{(1)}(y,\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} s^{(0)}(y,\boldsymbol{\beta})$ and $\mathbf{s}^{(2)}(y,\boldsymbol{\beta}) = \frac{\partial^2}{\partial \boldsymbol{\beta}^2} s^{(0)}(y,\boldsymbol{\beta})$ for all $y \in \mathcal{X}, \boldsymbol{\beta} \in B$.*

(e) *$\Sigma_{\boldsymbol{\beta}} = \int_{\mathcal{X}} \mathbf{v}(y,\boldsymbol{\beta}) s^{(0)}(y,\boldsymbol{\beta})\mu_0(y)dy$ is positive definite, where $\mathbf{v}(y,\boldsymbol{\beta}) = \frac{\mathbf{s}^{(2)}(y,\boldsymbol{\beta})}{s^{(0)}(y,\boldsymbol{\beta})} - \mathbf{e}^{\otimes 2}(y,\boldsymbol{\beta})$, $\mathbf{a}^{\otimes 2} := \mathbf{a}\mathbf{a}^T$, and $\mathbf{e} := \frac{\mathbf{s}^{(1)}}{s^{(0)}}$.*

(f) *There exists an $\eta > 0$ such that*

$$n^{-1/2} \sup_i |\mathbf{z}_i| \mathbf{1}(\boldsymbol{\beta}^T \mathbf{z}_i > -\eta|\mathbf{z}_i|) \xrightarrow{\text{p}} 0. \tag{21}$$

We now restate a modification of the theorems VII 2.1, 2.2 and 2.3 of Andersen et al. (1992).

Lemma 1. *If conditions (a), (b), (c'), (d) and (e) are satisfied, we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\text{d}} N(0, \Sigma_{\mathcal{X}}^{-1})$$

If further condition (c) is satisfied on a sub-interval $\mathcal{X}' \subseteq \mathcal{X}$, then $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is asymptotically independent of the process

$$W_{\mathcal{X}'}(x) = \int_{[0,x]\cap\mathcal{X}'} (d\hat{\text{M}}_0(y) - \mu_0(y)dy) + \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \int_{[0,x]\cap\mathcal{X}'} \mathbf{e}(y,\boldsymbol{\beta}_0)\mu_0(y)dy$$

and $W_{\mathcal{X}'}(\cdot)$ converges in distribution to the zero-mean Gaussian process $\int_{[0,x]\cap\mathcal{X}'} \sigma_W(x)dZ_W(x)$, where $Z_W(\cdot)$ follows standard Brownian motion, and $\sigma_W^2(x) = \mu_0(x)/s^{(0)}(x,\boldsymbol{\beta}_0)$.

The proof of the asymptotic normality of $\hat{\boldsymbol{\beta}}$ in Andersen et al. (1992) requires that the functions $\mathbf{e}(\cdot, \boldsymbol{\beta}_0)$ and $\mathbf{v}(\cdot, \boldsymbol{\beta}_0)$, together with the ratio $s^{(0)}(y, \boldsymbol{\beta})/s^{(0)}(y, \boldsymbol{\beta}_0)$ for $\boldsymbol{\beta} \in B$, be well-defined, hence the original condition (c) can be relaxed to our condition (c').

For our purposes, we are considering $\mathcal{X}' = \mathcal{X}_\epsilon = [x_l + \epsilon, x_u]$. Since we are considering cases in which random left truncation occurs, and the truncation variable x has a density on $[x_l, b_G]$ for some $b_G > 0$, we hence have $s^{(0)}(0, \boldsymbol{\beta}_0) = 0$, i.e. condition (c) is not satisfied on the whole of $\mathcal{X}$. However, this condition is satisfied on $\mathcal{X}_\epsilon$ for all $\epsilon > 0$, which is sufficient for our proof. The version of the asymptotic normality result for $\hat{\mathrm{M}}_0$ stated here follows from a straightforward modification of the proof of Theorem VII 2.3 in Andersen et al. (1992).

Proof of Theorem 1. We provide a proof for the case that G does not have a point mass at x_l. If it does, then $\int_{[0,\epsilon]\cap\mathcal{X}} \frac{\mu_0(y)}{s^{(0)}(y,\boldsymbol{\beta}_0)} dy < \infty$ for some $\epsilon > 0$, and one can apply the arguments of Keiding and Gill (1990) to obtain the uniform convergence of $\hat{\mathrm{M}}_0$ on the whole of $\mathcal{X}$, allowing for a much simplified proof.

Since $\mathcal{X}$ is a finite interval, without loss of generality let $\mathcal{X} = [0, 1]$. First we check the regularity conditions corresponding to Condition 1. We show that these are satisfied with $\mathbf{s}^{(m)}$, $m = 0, 1, 2$ as defined in the statement of the theorem.

Denote $S_x(y) = Pr(T_x > y - x|\mathbf{z} = 0) = {}_{y-x}p_x$. By the finiteness of the cumulative hazard, we get $S_x(y) \geq c$ uniformly in x, y for some $c > 0$.

$$
\begin{aligned}
s^{(0)}(y, \boldsymbol{\beta}) &= \mathbb{E}[\exp(\mathbf{z}'\boldsymbol{\beta})\mathbf{1}(x < y, T_x \wedge t > y - x)] \\
&= \int_0^y Pr(x + t > y)\mathbb{E}_{\mathbf{z}|x}[\exp(\mathbf{z}'\boldsymbol{\beta})(S_x(y))^{\exp(\mathbf{z}'\boldsymbol{\beta})}]dG(x) \\
&\geq \int_0^y k\mathbb{E}_{\mathbf{z}|x}[\exp(\mathbf{z}'\boldsymbol{\beta})c^{\exp(\mathbf{z}'\boldsymbol{\beta})}]dG(x) > 0.
\end{aligned}
$$

This lower bound increases in y, hence for $\epsilon > 0$, this implies that $s^{(0)}(\cdot, \boldsymbol{\beta}_0)$ is bounded below on $[\epsilon, 1] = \mathcal{X}_\epsilon$. Further, as $y \to 0$,

$$
\begin{aligned}
\frac{\|\mathbf{s}^{(m)}(y, \boldsymbol{\beta})\|}{s^{(0)}(y, \boldsymbol{\beta}')} &\geq \frac{p^{(m-1)_+}\mathbb{E}[\|\mathbf{z}\|^m \exp(\mathbf{z}'\boldsymbol{\beta})\mathbf{1}(x < y)]}{\mathbb{E}[\exp(\mathbf{z}'\boldsymbol{\beta}')\mathbf{1}(x < y)] - o(1)} \\
&= \frac{p^{(m-1)_+}\mathbb{E}[\|\mathbf{z}\|^m \exp(\mathbf{z}'\boldsymbol{\beta})|x < y]}{\mathbb{E}[\exp(\mathbf{z}'\boldsymbol{\beta}')|x < y]} + o(1) \\
&\to \frac{p^{(m-1)_+}\mathbb{E}[\|\mathbf{z}\|^m \exp(\mathbf{z}'\boldsymbol{\beta})|x = 0]}{\mathbb{E}[\exp(\mathbf{z}'\boldsymbol{\beta}')|x = 0]} \in (0, \infty).
\end{aligned}
$$

Using the above with $m = 0$, $\boldsymbol{\beta}' = \boldsymbol{\beta}_0$ gives the first part of condition (c'); whereas using $m = 1, 2$, $\boldsymbol{\beta} = \boldsymbol{\beta}' = \boldsymbol{\beta}_0$ gives the second part.

Now fix $\epsilon > 0$, and let $x \geq \epsilon$. Denote $\mathrm{M}_x(y) = (\mathrm{M}_0(y) - \mathrm{M}_0(x))_+ = \int_x^y dM_0(s)$. The cumulative hazard function corresponding to a person with covariate $\mathbf{z}_0$ is $\mathrm{M}_x(y; \mathbf{z}_0) = \lambda_0^T \mathrm{M}_x(y)$ where $\lambda_0^T = \exp(\mathbf{z}_0^T \boldsymbol{\beta}_0)$. Let $\hat{\mathrm{M}}_x(y; \mathbf{z}_0) = \hat{\lambda}_0 \hat{\mathrm{M}}_x(y)$ where $\hat{\lambda}_0 = \exp(\mathbf{z}_0^T \hat{\boldsymbol{\beta}})$; using the delta method with $g(\boldsymbol{\beta}, \mathrm{M}) = \exp(\mathbf{z}_0^T \boldsymbol{\beta}) \mathrm{M}$, we get

$$
\begin{aligned}
\nabla g(\boldsymbol{\beta}_0, \mathrm{M}) &= \lambda_0^T \begin{pmatrix} \mathbf{z}_0 \mathrm{M} \\ 1 \end{pmatrix} \\
\implies \sqrt{n}\left(\hat{\mathrm{M}}_x(y; \mathbf{z}_0) - \mathrm{M}_x(y; \mathbf{z}_0)\right) &\sim \lambda_0^T \left(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\mathbf{z}_0 \int_x^y \mu_0(s)ds + \sqrt{n}\int_x^y (d\hat{\mathrm{M}}_0(s) - \mu_0(s)ds)\right) \\
&\sim \lambda_0^T \left(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \int_x^y (\mathbf{z}_0 - \mathbf{e}(s, \boldsymbol{\beta}_0))\mu_0(s)ds + W_x(y)\right)
\end{aligned}
$$

where $W_x(y) := \int_x^y dW(s) = \int_x^y \sigma_W(s)dZ_W(s)$.

Define $\hat{S}_x(y;\mathbf{z}_0) = \exp(-\hat{\mathrm{M}}_x(y;\mathbf{z}_0))$ and $S_x(y;\mathbf{z}_0) = \exp(-M_x(y;\mathbf{z}_0))$. Applying the delta method yields

$$\sqrt{n}\left(\hat{S}_x(y;\mathbf{z}_0) - S_x(y;\mathbf{z}_0)\right) \sim -S_x(y;\mathbf{z}_0)\sqrt{n}\left(\hat{\mathrm{M}}_x(y;\mathbf{z}_0) - \mathrm{M}_x(y;\mathbf{z}_0)\right)$$

Finally define $\hat{e}_x^{(\hat{\lambda}_0)} = \int_x^\infty \hat{S}_x(y;\mathbf{z}_0)dy$ and $\mathring{e}_x^{T,(\lambda_0^T)} = \int_x^\infty S_x(y;\mathbf{z}_0)dy$. We obtain the asymptotic distribution of $\hat{e}_x^{(\hat{\lambda}_0)}$ via the functional delta method (see e.g. van der Vaart (1998), Ch. 26 for details); the functional $t(h) = \int h dy$ is Hadamard differentiable with derivative $t'(h) = \int h dy$ (this can also be motivated by the linearity of the integral). This hence gives

$$\begin{aligned}
\sqrt{n}(\hat{e}_x^{(\hat{\lambda}_0)} - \mathring{e}_x^{(\lambda_0^T)}) &\sim -\int_{y\geq x} S_x(y;\mathbf{z}_0)\sqrt{n}\left(\hat{\mathrm{M}}_x(y;\mathbf{z}_0) - \mathrm{M}_x(y;\mathbf{z}_0)\right)dy \\
&= -\int_{y\geq x}\int_{s\geq x} S_x(y;\mathbf{z}_0)\mathbf{1}(x\leq s\leq y)\sqrt{n}d(\hat{\mathrm{M}}_x(s;\mathbf{z}_0) - M_x(s;\mathbf{z}_0))dy \\
&= -\int_{s\geq x}\int_{y\geq s} S_x(y;\mathbf{z}_0)dy\sqrt{n}d(\hat{\mathrm{M}}_x(s;\mathbf{z}_0) - \mathrm{M}_x(s;\mathbf{z}_0)) \\
&= -\int_{s\geq x} S_x(s;\mathbf{z}_0)\mathring{e}_s^{T,(\lambda_0^T)}\sqrt{n}d(\hat{\mathrm{M}}_x(s;\mathbf{z}_0) - \mathrm{M}_x(s;\mathbf{z}_0))
\end{aligned}$$

where the use of Fubini's theorem is justified as the integral, being asymptotically normal, is bounded in probability.

Now define $EDLE_\epsilon = \frac{1}{n}\sum_{i:x_i>\epsilon}(\hat{e}_{x_i}^{(\hat{\lambda}_i)} - e_{x_i}^{T,(\lambda_i^T)})$ where $\hat{\lambda}_i = \exp(\mathbf{z}_i^T\hat{\boldsymbol{\beta}})$, $\lambda_i^T = \exp(\mathbf{z}_i^T\boldsymbol{\beta}_0)$. Similarly denote $\lambda^T = \exp(\mathbf{z}'\boldsymbol{\beta}_0)$, $\hat{\lambda} = \exp(\mathbf{z}'\hat{\boldsymbol{\beta}})$. We have

$$\begin{aligned}
\sqrt{n}EDLE_\epsilon &\sim \sqrt{n}\mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x>\epsilon)(\hat{e}_x^{(\hat{\lambda})} - \mathring{e}_x^{T,(\lambda^T)})\right] \\
&\sim \mathbb{E}_{x,\mathbf{z}}\left[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\mathbf{1}(x>\epsilon)\int_\epsilon^\infty \lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}(\mathbf{z} - \mathbf{e}(y,\boldsymbol{\beta}_0))\mu_0(y)dy\right] \\
&\quad + \mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x>\epsilon)\int_\epsilon^\infty \lambda\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}dW(y)\right] \\
&= \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\int_\epsilon^\infty \mathbb{E}_{x,\mathbf{z}}\left[\lambda\mathbf{1}(\epsilon<x<y)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}(\mathbf{z} - \mathbf{e}(y,\boldsymbol{\beta}_0))\right]\mu_0(y)dy \\
&\quad + \int_\epsilon^\infty \mathbb{E}_{x,\mathbf{z}}\left[\lambda\mathbf{1}(\epsilon<x<y)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}\right]dW(y)
\end{aligned}$$

We want to show that as $\epsilon\to 0$, $\sqrt{n}EDLE_\epsilon$ is asymptotically equivalent to $\sqrt{n}EDLE$, which we conjecture is asymptotically equivalent to

$$\begin{aligned}
\sqrt{n}EDLE &\sim \sqrt{n}\mathbb{E}_{x,\mathbf{z}}\left[\hat{e}_x^{(\hat{\lambda})} - \mathring{e}_x^{T,(\lambda^T)}\right] \\
&\sim \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\int_0^\infty \mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}(\mathbf{z} - \mathbf{e}(y,\boldsymbol{\beta}_0))\right]\mu_0(y)dy \\
&\quad + \int_0^\infty \mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}\right]dW(y)
\end{aligned}$$

First, we note that this conjectured limiting distribution is well-defined if its variance exists: this is given by $\mathbf{k}_{\boldsymbol{\beta}}'\Sigma_{\mathcal{X}}^{-1}\mathbf{k}_{\boldsymbol{\beta}} + \tau^2$ (refer to the statement of the theorem for the definitions of these quantities). Also, note that we can have $\tau^2 = \int_0^\infty\left(\mathbb{E}_{x,\mathbf{z}}\left[\lambda\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}_y^{T,(\lambda^T)}\right]\right)^2\frac{\mu_0(y)}{s^{(0)}(y,\boldsymbol{\beta}_0)}dy < \infty$ even if $\int_0^\epsilon\frac{\mu_0(y)}{s^{(0)}(y,\boldsymbol{\beta}_0)}dy = \infty$ for all $\epsilon > 0$.

To show that the difference between $\sqrt{n}EDLE_\epsilon$ and $\sqrt{n}EDLE$ is negligible, we note that this is given by

$$\begin{aligned}
\sqrt{n}(EDLE - EDLE_\epsilon) &= \frac{1}{\sqrt{n}}\sum_{x_i\leq\epsilon}(\hat{e}_{x_i}^{(\hat{\lambda}_i)} - e_{x_i}^{T,(\lambda_i^T)}) \\
&\sim \sqrt{n}\mathbb{E}_{x,\mathbf{z}}[\mathbf{1}(x\leq\epsilon)(\hat{e}_x^{(\hat{\lambda})} - e_x^{T,(\lambda^T)})]
\end{aligned}$$

Note that for $|\hat{e}_x^{(\hat{\lambda})} - \mathring{e}_x^{T,(\lambda^T)}| \leq x_u - x_l < \infty$, uniformly in x and $\mathbf{z}$. This gives

$$\sqrt{n}|EDLE - EDLE_\epsilon| \leq \sqrt{n}G(\epsilon)(x_u - x_l)$$

So that for ϵ_n such that $G(\epsilon_n) = o(n^{-1/2})$ this criteria is fulfilled and the proof is complete for the convergence in distribution.

For the consistency of the standard error, first note that the consistency of $\hat{\Sigma}_{\mathcal{X}}$ is given by Theorem VII 2.2 of Andersen et al. (1992). To simplify notation we let

$$\hat{w}_i(y) = \hat{\lambda}_i \hat{S}_{x_i}(y; \mathbf{z}_i)\hat{e}_y^{(\hat{\lambda}_i)}; \qquad w_i(y) = \lambda_i^T S_{x_i}(y; \mathbf{z}_i)\mathring{e}_y^{T,(\lambda_i^T)}; \qquad w(y) = \lambda^T S_x(y; \mathbf{z})\mathring{e}_y^{T,(\lambda^T)}$$

Hence we have

$$\begin{aligned}
&\|\hat{\mathbf{k}}_{\boldsymbol{\beta}} - \mathbf{k}_{\boldsymbol{\beta}}\| \\
&\leq \int_{\mathcal{X}} \left\| n^{-1} \sum_{i:x_i<y} \hat{w}_i(y) \left\{ \frac{(\mathbf{z}_i - \mathbf{E}(y, \hat{\boldsymbol{\beta}}))}{n^{-1}S^{(0)}(y, \hat{\boldsymbol{\beta}})} - \frac{(\mathbf{z}_i - \mathbf{e}(y, \hat{\boldsymbol{\beta}}))}{s^{(0)}(y, \hat{\boldsymbol{\beta}})} \right\} \right\| \frac{dN(y)}{n} \\
&+ \int_{\mathcal{X}} \left\| n^{-1} \sum_{i:x_i<y} \left\{ \frac{\hat{w}_i(y)(\mathbf{z}_i - \mathbf{e}(y, \hat{\boldsymbol{\beta}}))}{s^{(0)}(y, \hat{\boldsymbol{\beta}})} - \frac{w_i(y)(\mathbf{z}_i - \mathbf{e}(y, \boldsymbol{\beta}_0))}{s^{(0)}(y, \boldsymbol{\beta}_0)} \right\} \right\| \frac{dN(y)}{n} \\
&+ \int_{\mathcal{X}} \left\| n^{-1} \sum_{i:x_i<y} w_i(y)(\mathbf{z}_i - \mathbf{e}(y, \boldsymbol{\beta}_0)) - \mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x < y)w(y)(\mathbf{z} - \mathbf{e}(y, \boldsymbol{\beta}_0))\right] \right\| \frac{dN(y)}{ns^{(0)}(y, \boldsymbol{\beta}_0)} \\
&+ \left\| \int_{\mathcal{X}} \mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x < y)w(y)(\mathbf{z} - \mathbf{e}(y, \boldsymbol{\beta}_0))\right] \left(\frac{dN(y) - S^{(0)}(y, \boldsymbol{\beta}_0)\mu_0(y)dy}{ns^{(0)}(y, \boldsymbol{\beta}_0)} \right) \right\| \\
&+ \left\| \int_{\mathcal{X}} \mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x < y)w(y)(\mathbf{z} - \mathbf{e}(y, \boldsymbol{\beta}_0))\right] \left(\frac{S^{(0)}(y, \boldsymbol{\beta}_0)}{ns^{(0)}(y, \boldsymbol{\beta}_0)} - 1 \right) \mu_0(y)dy \right\|
\end{aligned}$$

We mimic the proof of the last part of Theorem VII 2.2 in Andersen et al. (1992). First note that $N(y)/n$ is bounded in probability; hence since the integrand of the first term vanishes by parts (a), (b) and (c) of condition 1, the integral as a whole vanishes as well. The same occurs with the second term; uniform consistency of the summands is given by their dependence on the underlying quantities $\hat{\boldsymbol{\beta}}$ and W, and since all the x_i's are larger than 0.

For the 3rd integral, we use the uniform law of large numbers of Rao (1963). It extends the law of large numbers to functions defined on a compact space: specifically, if $X_1, \ldots, X_n$ are i.i.d. random functions mapping $[0,1]$ to the real line, where $\mathbb{E}\sup_{t\in[0,1]} |X_1(t)| < \infty$, then

$$\sup_{t\in[0,1]} \left| \frac{1}{n} \sum_{i=1}^{n} X_i(t) - \mathbb{E}(X_1(t)) \right| \to 0 \text{ almost surely as } n \to \infty.$$

By the existence of $\mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}]$ and $\mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}\mathbf{z}]$, the boundedness of $\mathbf{e}(\cdot, \boldsymbol{\beta}_0)$, and since $S_x(y; \mathbf{z}) \leq 1$ for all $y \in \mathcal{X}$, $\mathbf{z} \in \mathbb{R}^p$, we first have

$$\mathbb{E}_{x,\mathbf{z}}[\sup_{y\in\mathcal{X}} w(y)\mathbf{e}(y, \boldsymbol{\beta}_0)] \leq k\mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}] \text{ for some } k > 0$$

$$\mathbb{E}_{x,\mathbf{z}}[\sup_{y\in\mathcal{X}} w(y)\mathbf{z}] \leq \mathbb{E}_{x,\mathbf{z}}[\lambda^T \mathring{e}_x^{T,(\lambda^T)}\mathbf{z}]$$

by the law of large numbers we get, uniformly in y,

$$\sup_{y\in\mathcal{X}} \left| n^{-1} \sum_{i:x_i<y} w_i(y)(\mathbf{z}_i - \mathbf{e}(y, \boldsymbol{\beta}_0)) - \mathbb{E}_{x,\mathbf{z}}[w(y)(\mathbf{z} - \mathbf{e}(y, \boldsymbol{\beta}_0)] \right| \overset{a.s.}{\to} 0$$

and hence since the integrand of the 3rd integral vanishes uniformly, so does the integral.

For the 4th integral, we use apply Lenglart's inequality in conjunction with counting processes: for a counting process N with compensator Λ such that $\sup_{x\in\mathcal{X}} \Lambda(x) < \infty$, then letting $M = N - \Lambda$ be the associated martingale, for an arbitrary function g,

$$Pr\left(\sup_{x\in\mathcal{X}} \| \int_0^x g(y)dM(y)\| > \eta\right) \leq \frac{\delta}{\eta^2} + Pr\left(\int_{\mathcal{X}} g^2(y)d\Lambda(y)\right).$$

In our case, the counting process N has the compensator Λ satisfying $d\Lambda(y) = S^{(0)}(y, \boldsymbol{\beta}_0)\mu_0(y)dy$. Letting

$$m(y) = \left\{\frac{\mathbb{E}_{x,\mathbf{z}}\left[\mathbf{1}(x<y)w(y)(\mathbf{z}-\mathbf{e}(y,\boldsymbol{\beta}_0))\right]}{s^{(0)}(y,\boldsymbol{\beta}_0)}\right\}$$

$$Pr\left(\left\|\int_{\mathcal{X}} m(y)\left(\frac{dM(y)}{n}\right)\right\| > \eta\right) \leq \frac{\delta}{\eta^2} + Pr\left(\frac{1}{n^2}\int_{\mathcal{X}} m^2(y)S^{(0)}(y,\boldsymbol{\beta}_0)\mu_0(y)dy > \delta\right)$$

By conditions (a), (b) and (c) of 1, we see that the latter probability goes to 0 for all $\delta > 0$, and hence so does the left-hand-side probability. Finally the same conditions imply the vanishing of the 5th integral as well.

Similar arguments give the consistency of $\hat{\tau}^2$ for τ^2.

□

Proof of Theorem 2. Similar arguments as in the proof of Theorem 1 give that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(\hat{e}^{(\hat{\lambda}_i)}_{x_i:x^*} - \mathring{e}^{T,(\lambda_i^T)}_{x_i:x^*}) \sim \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}^{T,(\lambda^T)}_{y:\overline{(x^*-y)|}}(\mathbf{z}-\mathbf{e}(y,\boldsymbol{\beta}_0))\right]\mu_0(y)dy$$
$$+\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\mathring{e}^{T,(\lambda^T)}_{y:\overline{(x^*-y)|}}\right]dW(y)$$

Hence it remains to find the contribution due to the imputation of the residual life,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(\hat{S}_{x_i}(x^*;\mathbf{z}_i)\mathring{e}^{(\lambda_i)}_{x^*} - S_{x_i}(x^*;\mathbf{z}_i)\mathring{e}^{T,(\lambda_i^T)}_{x^*}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\mathring{e}^{T,(\lambda_i^T)}_{x^*}(\hat{S}_{x_i}(x^*;\mathbf{z}_i) - S_{x_i}(x^*;\mathbf{z}_i)),$$

which is asymptotically equivalent to

$$\mathbb{E}_{x,\mathbf{z}}[\mathring{e}^{T,(\lambda^T)}_{x^*}\cdot\sqrt{n}\{\hat{S}_x(x^*;\mathbf{z}) - S_x(x^*;\mathbf{z})\}] \sim \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}[\mathbf{1}(x<y)\lambda^T\mathring{e}^{T,(\lambda^T)}_{x^*}S_x(x^*;\mathbf{z})(\mathbf{z}-e(y,\boldsymbol{\beta}_0))]\mu_0(y)dy$$
$$+\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}[\mathbf{1}(x<y)\lambda^T S_x(x^*;\mathbf{z})\mathring{e}^{T,(\lambda^T)}_{x^*}]dW(y)$$

Combining gives

$$\sqrt{n}EDLE \sim \sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\underbrace{(\mathring{e}^{T,(\lambda^T)}_{y:\overline{(x^*-y)|}} + S_y(x^*;\mathbf{z})\mathring{e}^{T,(\lambda^T)}_{x^*})}_{\mathring{e}^{T,(\lambda^T)}_y}(\mathbf{z}-\mathbf{e}(y,\boldsymbol{\beta}_0))\right]\mu_0(y)dy$$

$$+\int_{\mathcal{X}^*}\mathbb{E}_{x,\mathbf{z}}\left[\lambda^T\mathbf{1}(y>x)S_x(y;\mathbf{z})\underbrace{(\mathring{e}^{T,(\lambda^T)}_{y:\overline{(x^*-y)|}} + S_y(x^*;\mathbf{z})\mathring{e}^{T,(\lambda^T)}_{x^*})}_{\mathring{e}^{T,(\lambda^T)}_y}\right]dW(y)$$

□