Evaluating Life Expectancy Evaluations[1]

Daniel Bauer, Michael V. Fasano, Jochen Russ, and Nan Zhu.[2]

Abstract

The quality of life expectancy estimates is one key consideration for an investor in life settlements. The predominant metric for assessing this quality is the so-called *A-To-E Ratio*, which relies on a comparison of the actual to the predicted number of deaths. In this paper, we explain key issues with this metric: In the short run, it is subject to estimation uncertainty even for a moderately-sized portfolio; and, in the long run, it converges to 100% even if the underwriting is systematically biased. As an alternative, we propose and discuss a set of new metrics based on the *Difference in (Temporary) Life Expectancies*. We examine the underwriting quality of a leading US life expectancy provider based on this new methodology.

Keywords: Life Settlement; Life Insurance; Investment; Secondary Market; Life Expectancy Evaluation; A/E Ratio; Difference in Temporary Life Expectancies

---

[1] Parts of this paper are taken from a newsletter article by the first and third author based on a presentation at the Fasano 9th Annual Life Settlement & Longevity Conference, and parts are taken from an old version of the paper "Adverse Selection in Secondary Insurance Markets: Evidence from the Life Settlement Market" by the first, third, and fourth author.

[2] Daniel Bauer is the Robert W. Batten Chair in Actuarial Science in the Department of Risk Management and Insurance at Georgia State University (dbauer@gsu.edu). Michael V. Fasano is President of Fasano Associates, a medical underwriting firm and life expectancy provider (mfasano@fasanoassociates.com). Jochen Russ is managing partner of the Institute for Finance and Actuarial Science, an actuarial consulting firm based in Ulm, Germany and an adjunct professor for actuarial science at Ulm University (j.russ@ifa-ulm.de). Nan Zhu (corresponding author) is an Assistant Professor of Risk Management in the Risk Management Department at Pennsylvania State University (nanzhu@psu.edu, phone: 814-863-8666, fax: 814-865-6284).

1 Introduction

Life expectancy (LE) providers specialize in compiling individualized mortality forecasts within the underwriting process in primary and secondary life insurance markets. Particularly for investors in secondary market transactions, so-called *life settlements*, the policyholder's (unknown) remaining life expectancy is the major driver for the value of a policy—and, hence, for the price at which such a policy is traded. Therefore, the quality of LE-estimates used in the valuation of such a transaction is one of the most important aspects of the investment decision. So far, the predominant metric for assessing this quality is an ex-post comparison of the *Actual* number of deaths to the predicted or *Expected* number of deaths, the so-called *A/E (A-To-E) ratio* (Actuarial Standards Board, 2013). If the actual number is close to the expected number, then the ratio of actual to expected deaths obviously will be close to 100%—which typically is interpreted as an indication of a high quality of medical underwriting.

In this paper, we commence by explaining why such an interpretation can be deceiving. Using an illustrative example, we first show that relatively shortly after the point of underwriting, relatively few deaths occur, so that the metric is subject to estimation uncertainty even for moderately sized portfolios. More importantly, we show that over time, the A/E ratio becomes increasingly less meaningful: Since eventually all lives that were expected to die (E) will indeed die (A), the A/E ratio will automatically be pulled towards 100%. Since existing portfolios are frequently made up of policies that were underwritten relatively recently and policies that were underwritten less recently, the extent to which the latter "bias" contributes to the overall underwriting evaluation is generally unclear. Thus, in many situations, the metric may not be suitable for revealing the true quality of the LE forecasts to analysts and investors.

As an alternative, we propose and analyze a set of new methodologies for measuring actual to expected performance via the *difference in life expectancies* (*DLE*), the *difference in temporary life expectancies* (*DTLE*), the *normalized difference in temporary life expectancies* (*NDTLE*), and the *indirect difference in life expectancies* (*IDLE*). Here the DLE is simply the difference in realized and estimated life expectancy, calculated as the average of realized lifetimes less the LE-estimates across a portfolio. We argue that while the first metric, DLE, has the most desirable properties in that it directly corresponds to the quantity of interest, its calculation is not possible until all individuals in the portfolio have died. The DTLE, on the other hand, is the difference of the temporary lifetime realized until the time of observation—that is the minimum of the lifetime and the observation period—less its expected value under the provider's estimates, which in the actuarial literature is usually referred to as *temporary life expectancy* (Bowers et al., 1997). Thus, an estimate for the DTLE metric can easily be obtained as the average of the differences in temporary lifetimes and temporary LE-estimates.

For a given portfolio, the DTLE converges to the DLE over time as the temporary life expectancy converges to life expectancy. Hence, unlike the A/E ratio, the informative value of the DTLE will improve as the portfolio ages and eventually converge to the metric of interest. Therefore, the DTLE overcomes the main issue of the conventional A/E metric.

Furthermore, the standard errors and thus the uncertainty of the estimate are relatively small even early on when only few people have died. This latter aspect is associated with a drawback in the interpretation, however, since the information about deaths is sparse and is only incorporated over time. To elaborate, while a DTLE in the long run will be close to the DLE and so a DTLE of, say, 0.5 is easily interpreted, the significance of a DTLE of 0.1 early on is difficult to assess, since it is not clear if this value will result in a final value of, say, 0.5 years or 2 years.

As a metric that combines both advantages, easy interpretability early on and meaningful results as the time passes, we further introduce *normalized* versions of the DTLE (NDTLE). This normalization can take the form of dividing the DTLE by the hypothetical DTLE that would correspond to a certain final DLE (e.g., one year in our examples). Then, an NDTLE of around +/- 100% indicates that the underwriter's estimate will be one year too short (aggressive) or one year too long (conservative), respectively. Similarly, an NDTLE of +/- 200% indicates that the LE-estimates are on average roughly two years off in the aggressive or conservative direction. Alternatively, instead of normalizing by an absolute, fixed difference in eventual life expectancies, it is also possible to normalize using a relative percentage by which the final realized LE differs from its estimate. For instance, if one uses 20%, an NDTLE of around +/- 100% indicates that the underwriter's estimate will be on average 20% too short or 20% too long, respectively. This may have advantages in the context of heterogeneous portfolios, because clearly a one-year deviation has different implications depending on the length of the LE-estimate. Beyond normalizing, we also solve for the absolute or relative adjustment itself as an indirect quality metric, the indirect difference in life expectancies (IDLE). More precisely, we find the eventual deviation in DLE (in either absolute or relative sense) that would be consistent with a current NDTLE of 100%. Using our proposed new metrics, the results are easy to interpret, and properties of each metric—including asymptotic distributions and thus confidence bands—are easy to derive. In particular, we show that it will converge to the actual error of the estimated LE in the long run. This property should make it preferable to analysts and investors for assessing the quality of LE-estimates.

We apply the above new metrics using data from a large US LE-provider, Fasano Associates, in order to analyze its underwriting performance from beginning-of-year 2001 through end-of-year 2013. Our results show that underwriting has been spot on for the more recent time period: For lives that were underwritten from year 2006, the DTLE, NDTLE, and IDLE are all very close to zero. For the entire underwriting portfolio, the metrics are slightly higher: The DTLE is less than two months, which further corresponds to an estimated average deviation of the eventual DLE at around eight months or relatively 8%. We note that this is to a great extent caused by more aggressive underwriting in the early 2000s as evidenced by our time-development analyses. The metrics for the full portfolio are jointly influenced by both very good recent underwriting as well as lives too aggressively underwritten in the early years.

Related Literature and Organization of the Paper

As pointed out by Seligman and Kahn (1980), the comparison of actual and expected results has a long history in retrospective actuarial studies. The authors derive hypothesis tests for the correctness of the assumptions in calculating the expected values, by deriving the (asymptotic) distribution of the numerator in various instances. Rhodes and Freitas (2004) also derive approximate confidence bands for A/E ratios in the context of a Poisson model. We contribute to this line of literature by proposing a new set of metrics and analyzing asymptotic properties, including the calculation of asymptotic confidence intervals.

Various authors indicate the significance of medical underwriting results for life settlement investors (Bhuyan, 2009; Aspinwall, Chaplin, and Venn, 2009). In particular, Gatzert (2010) points to the underwriters' performance as a possible explanation of the gap between expected and realized investment returns in the life settlement market, and Januário and Naik (2014) provide empirical evidence on the relevance of "life expectancy estimation risk" for life settlement return (see also Braun, Gatzert, and Schmeiser (2012); Davó, Resco, and Barroso (2013); Giaccotto, Golec, and Schmutz (2015) for studies on life settlement investment performance). While there are general difficulties associated with composing and evaluating underwriting performance (Qureshi and Fasano, 2010), the specific problems associated with the bias of the A/E ratio may also lead to a misevaluation of underwriting performance by investors, and thus may exacerbate the issue of "life expectancy estimation risk." Hence, our paper contributes by providing investors alternative—and as we argue superior—metrics, and by informing the marketplace about the results of one large underwriter based on this new approach.

The paper is organized as follows. In Section 2, we first provide analyses of the A/E ratio documenting the indicated deficiencies. In Section 3, we then introduce the novel DLE/DTLE/NDTLE/IDLE metrics as alternatives, discuss their properties, and present illustrative calculations. In Section 4, we calculate empirical results using the Fasano data. Finally, we present our conclusions in Section 5.

2 A/E Ratios with Illustration

A/E ratios are commonly used in actuarial practice to assess the quality of LE forecasts by tracking the ratio of the actual number of deaths to the expected number of deaths over time (Seligman and Kahn, 1980). While in practice A/E ratios are typically reported based on the development over calendar years for a given portfolio, we focus here on the development over years since underwriting in order to better illustrate the drawbacks of the metric. More precisely, for each individual i in a portfolio of size N, we denote by $\tau_x^{(i)}$ the individual's realized remaining lifetime and by ${}_tq_x^{(i)}$ the individually estimated t-year mortality probability as seen from the time of underwriting, where in line with standard actuarial notation x is the individual's age (at the time of underwriting). The A/E ratio for the portfolio of all individuals that have been underwritten for at least t years is hence:

$$\frac{A}{E} = \frac{\sum_{i=1}^{N} 1_{\left\{\tau_x^{(i)} < t\right\}}}{\sum_{i=1}^{N} {}_t q_x^{(i)}},$$

where $1_{\left\{\tau_x^{(i)} < t\right\}}$ is the indicator function that takes the value 1 if individual i died before time t and 0 otherwise. Hence, A is simply the number of people who have actually died before time t, whereas E is the expected number of deaths before time t under the assumption that the estimated mortality probabilities (corresponding to the estimated LEs) are correct.

To illustrate the deficiencies of the A/E ratio as a measure of the quality for life expectancy estimates, we consider a very simple hypothetical example. We assume an LE portfolio composed of 500 identical non-smoking male individuals aged 75 who have actual mortality probabilities twice as high as the probabilities from the standard VBT 2008 mortality table.[3] Moreover, we contemplate two biased LE providers with the first (A) being too aggressive in assessing LEs and the second provider (C) being too conservative. More precisely, we assume that each LE provided by (A) is two years too short whereas provider (C)'s LEs are two years too long. We further assume that the deviations originate only from an erroneous assessment of the "frailty factor" to be multiplied on the mortality probabilities from the VBT table. That is, we assume that there are no systematic deviations between the mortality table used by the underwriter and the "correct" table.

The solid lines in Figure 1 show the corresponding development of A/E ratios for both the aggressive LE provider (A) and the conservative LE provider (C) if the 500 people die exactly according to their mortality probabilities (i.e. without any random fluctuation). Due to random fluctuation, the actual number of deaths will of course not evolve exactly according to the mortality probabilities. As a result, A/E ratios in reality will also deviate from the solid lines. The dashed lines therefore show the 95% confidence intervals of the ratios. There are a number of contributions in the actuarial and statistical literature that provide analyses of the A/E ratio, including formulas for approximating asymptotic confidence intervals (see e.g. Rhodes and Freitas (2004) and references therein). However, here we rely on Monte Carlo estimates for the 2.5% and 97.5% percentiles of the A/E ratios for each year, based on 100,000 simulations of the random times of deaths in the portfolio as they should provide accurate estimates for the finite-sample intervals.

There are two striking observations. First, the confidence intervals of the A/E ratios in the early years are very wide. In fact, despite the considerable difference of four years in their LE assessment, each deviating two years from the correct values, the 95% confidence intervals for the two providers overlap for the first three years since underwriting. The intuition for this observation is straightforward: The probability of dying immediately after underwriting is relatively low, so the denominator in the A/E ratio is very small. Therefore, if some individuals die—or fail to die—early on, the impact on the ratio will be considerable. This indicates that the A/E ratio has to be interpreted with caution for *young* portfolios.

[3] We choose a multiplier ("frailty factor") of 200% since the life settlement market typically targets policyholders with below-average health status.

Second, the A/E ratios for both the aggressive and the conservative LE provider are approaching 100%. For instance, after less than 20 years, the A/E ratios are well within the range between 90% and 110% for both providers with a high probability—even though both providers were systematically and significantly wrong in their original estimates. Again, the intuition is straightforward: Both, the actual number of deaths and the expected number of deaths will necessarily approach the number of people in the portfolio since, eventually, all individuals will die. Hence, both A/E ratios are "pulled" towards 100% in the long run, even if one provider was too aggressive and the other too conservative. This indicates that the A/E ratio has to be interpreted with caution for *old* portfolios.

Of course, as the size of the portfolio N increases, the statistical uncertainty decreases and the results will be more accurate early on. For instance, in Figure 2, we plot the development of the A/E ratio and associated confidence intervals similarly as in Figure 1 but for a portfolio size of N=50,000, which roughly corresponds to our empirical application in Section 4 (the Fasano portfolio). We find that the widths of the confidence intervals for both A/E ratios become much tighter. Importantly, however, the long run bias is of systematic nature and is therefore not alleviated by the larger sample size.

Thus, we conclude that the A/E ratio is not meaningful to assess underwriting quality in the long run. Typical portfolios include both, lives that were underwritten recently as well as lives underwritten further in the past. Therefore, it is difficult to assess the influence of this effect. Additionally, shortly after underwriting, there is considerable estimation uncertainty for moderately sized portfolios. While this aspect may not be material for the full portfolio of a large LE provider, A/E analyses are also frequently used for moderately sized portfolios (e.g. an investor's portfolio or an LE-provider's sub-portfolio of, say, only males above age 75 with cancer). In such cases, both aspects may be of relevance.

3 DTLE Methodology with Illustration

To determine a metric that overcomes some of the deficiencies of the A/E ratio, first assume that we knew the time of death for each individual in a given portfolio of LEs. In this case we could calculate the difference between the realized lifetime and the given LE-estimate for each life. For a given individual, of course this would ex ante be a random variable due to the intrinsic randomness of the remaining lifetime. In other words, even if all LE-estimates were exactly accurate, the realized lifetime will be larger than its expected value for some whereas it will be less than the life expectancy for others. Under the hypothesis that the LE-estimates are accurate, the average, however, should be close to zero. This is exactly the intuition of the *law of large numbers*, which states that the average of (mean-zero) random variables with bounded variance will converge to zero. On the other hand, if the average of these differences were positive, that would be a strong indication that the estimated LEs were too short on average—especially with a sufficiently large portfolio—whereas a negative average would indicate that estimated LEs were too long on average. We refer to the average of these differences as the *difference in life expectancies* (DLE).

However, in relevant situations, not all individuals will have died when one wants to assess the quality of an LE-underwriter. In particular, we can only observe times of death that occurred before the present time. For all other individuals we only know that they are still alive at this cut-off date but we do not know when they will die. To deal with this so-called *right-censoring* issue, we rely on an alternative concept from actuarial science and demographic research. More precisely, denote by $\bar{\tau}_x^{(i)} = \min(\tau_x^{(i)}, t_i)$ the *realized partial lifetime*, which is the right-censored version of the time of death, i.e. the number of years lived between time of underwriting and the cut-off date. Here, t_i is the (given) difference between the cut-off date and the time of underwriting for individual i. Its expected value, the so-called *temporary life expectancy*, is then given by the integral of the survival probabilities until time t_i (Bowers et al., 1997):

$$e_{t_i}^{(i)} = E\left[\bar{\tau}_x^{(i)}\right] = \int_0^{t_i} {}_tP_x^{(i)} dt,$$

where ${}_tP_x^{(i)}$ denotes the (unknown) "true" survival probabilities for individual i. Furthermore, we denote the estimated survival probabilities for individual i supplied by the LE provider by ${}_tp_x^{(i)} = 1 - {}_tq_x^{(i)}$, and denote the corresponding estimated temporary life expectancy by:

$$\hat{e}_{t_i}^{(i)} = \int_0^{t_i} {}_tp_x^{(i)} dt.$$

Then, under the hypothesis that the estimated lifetime distributions are accurate, the difference in realized partial lifetime and estimated temporary life expectancy for individual i, $\bar{\tau}_x^{(i)} - \hat{e}_{t_i}^{(i)}$, is a zero-mean random variable with bounded variance. Thus, the average:

$$DTLE_N = \frac{1}{N}\sum_{i=1}^{N}\left[\bar{\tau}_x^{(i)} - \hat{e}_{t_i}^{(i)}\right],$$

which we will refer to as the *difference in temporary life expectancies* (DTLE), again converges to zero and is again asymptotically normally distributed according to Lyapunov's central limit theorem (Billingsley, 1995):

$$\frac{DTLE_N}{\frac{1}{N}S_N^{TR}} = \frac{1}{S_N^{TR}}\sum_{i=1}^{N}[\bar{\tau}_x^{(i)} - \hat{e}_{t_i}^{(i)}] \xrightarrow{d} N(0,1),$$

where $(S_N^{TR})^2 = \sum_{i=1}^{N} Var[\bar{\tau}^{(i)}] \approx \sum_{i=1}^{N}\left(\bar{\tau}_x^{(i)} - \hat{e}_{t_i}^{(i)}\right)^2$. Clearly, as time passes by and all the t_i increase, the difference in temporary life expectancy will converge to DLE. Since, as described above, the latter is the "actual" metric we are interested in—ignoring estimation uncertainty—the long-run value of the DTLE will exactly reflect the desired information.

To illustrate, we perform the DTLE analysis using the same hypothetical portfolio and assumptions of the two biased LE providers (A) and (C) as described in the previous section. Figure 3 shows how the DTLEs evolve over years since underwriting. Again, there are two striking observations. First, in the short run, the DTLE does not show a large deviation for either provider, and the confidence bands around the central estimates are also very narrow. While this may look like a deficiency, as indicated in the A/E analysis, the early deaths provide little usable information in view of the provider's quality. The point estimate for the A/E ratio ignores this and can take on values that heavily deviate from 100%; thus their statistical credibility is limited. The DTLE, on the other hand, reflects that there is little informational content in the early observations and is close to zero. If it is significantly different from zero, however, the result will be credible and likely reflect the provider's quality.

Second, as soon as there is sufficient statistical content to reflect the underwriting quality, the DTLE will diverge from zero and converge to the "true" DLE value. In particular, after approximately 20 years, the central estimates for the metric correctly identify that both LE providers were off by exactly two years in opposite directions. There exists estimation uncertainty, but this is unavoidable since it is observationally equivalent for a provider to be systematically biased or to have "bad luck" in the sense that the finite sample did not evolve according to the "true" distribution.

Similar to the previous section, the confidence intervals will be considerably tighter for large portfolios. Figure 4 again shows an analogous illustration for a portfolio size of *N=50,000*. It is worth noting that here, however, the portfolio size does not only impact the estimation early on but also in the long run. More precisely, it is unlikely that for large portfolios, "good" or "bad" luck will have a considerable impact for the eventual DTLE result. Hence, in the medium to long run and for sizable portfolios, the DTLE provides meaningful and easy-to-interpret results for the underwriting quality of an LE provider.

Although, as discussed, the small level of the DTLE in the early period after underwriting is caused by the absence of useful information since there are a limited number of deaths, its interpretation is difficult. For instance, while in our example a value of DTLE close to 2 after about 20 years is easily interpreted, the significance of a DTLE of roughly 0.2 after 5 years or even of roughly 0.5 after 10 years is difficult to assess. As a solution that combines both attributes, stability in the medium to long run and an easy interpretation also in early years after underwriting, we introduce the *normalized difference in temporary life expectancies* (NDTLE). Within the calculation of the NDTLE, the DTLE is divided (normalized) by a hypothetical benchmark DTLE that would correspond to a certain (absolute or relative) deviation in the final DLE. Thus, a ratio of 100% will indicate that the DTLE is in line with the assumed deviation in the final DLE.

More precisely, based on the survival probabilities from the underwriter's estimate ${}_tp_x^{(i)}$, we can derive a hypothetical level of ${}_tP_x^{(i),*}$ that would result in a certain imposed DLE. For instance, we could determine the ${}_tP_x^{(i),*}$ such that the DLE amounts to *d* years or *p* percent of the underwriter's original estimate if the individuals die according to the distribution implied by ${}_tP_x^{(i),*}$. Then we can calculate the benchmark DTLE ($\mathrm{DTLE_d}$ respectively $\mathrm{DTLE_p}$) based on this hypothetical distribution:

$$\mathrm{DTLE}_{d/p} = \frac{1}{N}\sum_{i=1}^{N}\left[\int_0^{t_i} {}_{t}P_x^{(i),*}dt - \hat{e}_{t_i}^{(i)}\right].$$

Now we define the $NDTLE_d$ and $NDTLE_p$ as DTLE/ $DTLE_d$ and DTLE/ $DTLE_p$, respectively, where the numerator in both formulas is the one calculated from the realized deaths. For instance, for *d=2* years, $DTLE_d$ and $DTLE_{-d}$ will correspond to the solid lines in Figures 3 and 4. As a consequence, $NDTLE_d$ would measure the *relative* deviation of the observed DTLE from these solid lines. Importantly, the NDTLE is simply a (deterministically) scaled version of the DTLE, so the transformation will not affect on the statistical results for the DTLE other than including a constant factor to each summand.

Figure 5 shows the evolution of the NDTLE for *d=1* year for our illustrative example. The central estimates of the NDTLE for both the aggressive (A) and the conservative (C) LE provider are now relatively flat. In particular, they are approximately 200% and -200% for A and C, respectively, since the providers' estimates are "twice as bad" as the absolute deviation (d) of 1 year used for the normalization. Hence, ignoring statistical uncertainty, an observer can conclude very early on that the realized difference will probably result in a final DLE (or in a DTLE in late years) of approximately two years.

Due to the normalization the early results get inflated so that, as for the A/E ratio, the credibility is low and corresponding standard errors are large. Hence, there exists some kind of "uncertainty principle" in that an easy-to-interpret early warning sign will necessarily have a large error associated with it. However, as with the A/E ratio, the error becomes less pronounced for large portfolio sizes *N* and, unlike the A/E ratio, results become *both,* credible and meaningful as time passes.

The central estimates in Figure 5 are not exactly 200% and -200% because of inherent nonlinearities. The DLE for the conservative provider (C) is -2 in our example and the normalization is based on a DLE of 1, so that NDTLE will eventually converge to -2/1 = -200%. However, after, say, 2 years $DTLE_{d=1\ year}$ is not necessarily half of the absolute value of C's DTLE, and indeed Figure 5 indicates that the ratio is between 1/2 and 1/3. To obtain this *exact* interpretation, it is also possible to define a metric *indirectly*. That is we can determine the final absolute deviation *d* (or, alternatively, the final *relative* deviation *p*) in DLE that would result in a current NDTLE of 100%. Formally, we determine the value of *d* (or *p*) such that $NDTLE_{p/d}$ = DTLE/$DTLE_{p/d}$ = 100%. We refer to the respective solution *d* (or *p*) as the *indirect difference in life expectancies* (IDLE).

Figure 6 plots the IDLE for our hypothetical example. Of course, now the central estimates for the aggressive (A) and conservative (C) LE providers are *exactly* 200% and -200%, respectively, since these are the levels of the eventual DLEs we assume in setting up the calculation.[4] Further, because of the monotonic transformation the confidence bands of IDLE are directly calculated as solutions of *d* (or *p*) that would result in an NDTLE being the same as the associated confidence bands of DTLE,. We observe from the figure that the confidence bands are again wide initially—due to the afore-mentioned "uncertainty principle"—but will

[4] To clarify, the central estimates will be *exactly* ±200% in the limit. Since here, the estimate is derived based on 100,000 repetitions of a Monte Carlo analysis, some estimation uncertainty remains.

narrow over time. In contrast to the NDTLE, however, the width is now relatively similar for both providers A and C.

4. Analyses on Fasano Associates Data

In this section, we assess the quality of the LE-estimates for the portfolio of a concrete LE-provider, Fasano Associates (Fasano), using the introduced DTLE/NDTLE/IDLE metrics. Our data consist of 53,947 distinct life expectancy evaluations underwritten by Fasano between beginning-of-year 2001 and end-of-year 2013.[5] For each record, we are given an LE-estimate that is the result of the underwriting process. Beyond these LE-estimates, we are also given realized times of death for the individuals that had died before January 1st 2015.

Two practical issues arise before specific analyses can be performed. First, we need to establish reasonable assumptions on the IBNR (*Incurred But Not Reported*) rate, since LE providers are typically not informed about an underwritten individual's death and have to rely on public data that are often inaccurate, incomplete, or delayed (Behan, 2012). In this paper we use a constant annual IBNR rate of 7%, which was recommended by a number of LE underwriters based on Social Security Master Death File data. Second, we need to fix a consistent evaluation methodology for all LEs in the portfolio, which of course were furnished at different points in time.[6] Here, we always use the original LE-estimates (i.e. we do not try to adjust for the fact that under current methodology, different LE-estimates would be given for some individuals) and we use Fasano's most current proprietary mortality table to generate mortality distributions around such point estimates. More precisely, we solve for a mortality multiplier (frailty factor) that has to be applied to this proprietary table in order to match the LE-estimate that was originally provided by Fasano. We use the resulting mortality probabilities in our analyses.

Table 1 shows the metrics of interest for the entire Fasano portfolio as of January 1st 2015, together with their 95% confidence intervals that are determined based on the normal approximation discussed in Section 3. From the table we first observe that the DTLE is positive and at a level of less than two months ($0.1442 \times 12 = 1.73$). While a deviation in this range on average certainly may appear innocuous on its own, we recall from the previous illustrations that the DTLE is always small initially because of the limited number of deaths. Since the entire portfolio we study here contains LE-estimates for lives that were underwritten both recently and not so recently, the interpretation of this number is not immediately clear. Therefore, we further calculate the NDTLE for two choices of the normalization: By assuming the eventual life expectancy exceeds the estimate for each life by an absolute d=1 year or by a relative p=20% of the initial estimate. Finally, we solve for the IDLE considering both the absolute and relative deviations. The results for NDTLE and IDLE are similar: For the entire portfolio, the LE-estimates were on average too short by approximately 8 months or a relative deviation of 8%, suggesting a rather good quality of LE estimates.

[5] We use the earliest observation in the analyses when the same individual has been underwritten several times.
[6] For instance, the Actuarial Standards Board (2013) recommends the so-called adjusted-to-current method, although interpretations of and opinions on this approach differ in the industry.

It is worth pointing out the considered LE provider—and more generally the industry—has been adapting its underwriting technology over the past decade. Therefore, the overall deviation in LE-estimates might be heavily attributed to an underwriting that was more aggressive in early years, but does not represent the current underwriting performance. Indeed, differences for lives underwritten early on will disproportionally impact the absolute versions of our metrics over the years, simply because they are larger in absolute value because of the longer history. Thus, we also separate the entire portfolio into two sub-portfolios based on the time of underwriting (before and after January 1st 2006), and calculate the same metrics for each sub-portfolio (also displayed in Table 1). We observe from the results that for lives that were underwritten more than a decade ago, the average bias in the underwriting is more pronounced (approximately one year, or a relative deviation of 13.5%), whereas for the more recent sub-portfolio, the underwriting is fairly accurate: Not only the point estimates for all metrics are very close to zero, their respective confidence intervals all contain the zero point.

Figure 7 further shows the development of the DTLE for the Fasano data from two distinct perspectives: The left-hand panel shows the evolution of the DTLE and its confidence band at the end of each calendar year from year 2001 to year 2014, using all individuals that had been underwritten by each of the respective cutoff date. The values at the end of year 2014 are the ones in Table 1. We observe that the figure exhibits a "smirk" shape. The DTLE for early years is (significantly) positive although the value is relatively low. Again, the latter is not surprising since the DTLE is naturally small over the early years. However, the observation that it is significant and positive corroborates that underwriting has been more aggressive for these early lives. While the DTLE first decreases to levels close to zero, it eventually increases to slightly below two months, a level that is disproportionally influenced by the early LE estimates as explained above.

To better illustrate the underwriting performance over time, as within our illustrations in earlier sections, we also plot the DTLE results using years since underwriting on the *x*-axis in the right-hand panel of Figure 7. As a consequence, the underlying dataset differs for each year (t). More precisely, the results for 1 year include all available data until end-of-year 2013—since we have reported deaths until beginning-of-year 2015 so that we observed all these lives at least for one full year. The results for, say, 10 years, however, include only the portion of the data that have been observed for more than 10 years, i.e. lives underwritten before beginning-of-year 2005. In particular, the results for low values of t include the most recent underwriting data and, thus, are the most relevant estimates for assessing the underwriter's current quality, whereas the results for high values of t are solely based on lives underwritten in earlier years and hence measure the quality of the underwriting methodology used at that time. The DTLE is close to 0 for low values of t, but then increases to levels close to 1 year for t=12. The latter corroborates insights from Table 1 and the left-hand panel of Figure 7 that early underwriting (around 2001-2002) was more aggressive, although now the level (DTLE$\approx$1) is more accurate. The finding that results for low values of t are close to zero confirms that underwriting has improved over the years, although as discussed the DTLE is difficult to interpret in this range.

Figures 8 and 9 depict the development of absolute (Figure 8) and relative (Figure 9) NDTLE over both, calendar year (left panel in each figure) and years since underwriting (right panel in each figure). In particular, Figure 8 normalizes by using a benchmark DTLE that would correspond to a deviation in the final DLE of d=1 year; and Figure 9 normalizes using a benchmark DTLE that would correspond to a relative deviation in the final DLE of p=20% of the initial estimates. From both figures, the early calendar-years in the left-hand panels and the large number of years-since-underwriting (t) in the right-hand panels confirm that LE-estimates were more aggressive in the early 2000s, with estimates being a little over 1 year or a little under 20% too short on average. However, results that include more recent underwriting experience show that this is no longer the case. Indeed, results for t=1—which as discussed include the entire portfolio experience—demonstrate the DTLE is close to zero, whereas the results for t from 2 to 5 suggest that the underwriting has even been slightly on the conservative side in some periods.

Figures 10 and 11, finally, depict the development of absolute IDLE (Figure 10) and relative IDLE (Figure 11) over both, calendar year (left panel in each figure) and years since underwriting (right panel in each figure). Hence, one can interpret the figures as today's estimate for the (absolute or percentage) deviation in the eventual life expectancy that can only be observed in the future, when all insured will have deceased. In other words, under the assumption that the (today unknown) DLE will turn out to be this value in the future, the expected DTLE today coincides with the observed value for DTLE—or the associated NDTLE using this deviation will be 100%. In line with Figures 8 and 9, this deviation is close to zero and even slightly negative for years since underwriting between 1 and 6, which include the lives underwritten in this decade. However, also in line with the previous discussion, the percentage of the increases amounts to roughly +10% for t=9—meaning that on average estimates compiled pre-2006 were approximately 10% too short.

5. Conclusion

The quality of life expectancy estimates is of utmost importance for all participants in the life settlements market. But the question how to measure underwriting quality is also of certain relevance in primary insurance markets. In order to assess the quality of LE-estimates, typically so-called A/E ratios are being used. In the present paper, we explain that this metric has substantial shortcomings, both, for "young" and "old" portfolios. In particular, an A/E ratio for an old portfolio is artificially "pulled" towards 100% even if LE-estimates were systematically and significantly too short or too long. Therefore, under certain circumstances this metric can be misleading or even meaningless.

In order to overcome these shortcomings, we propose several alternatives. The number that is directly linked to the quality of the LE-estimates is the difference in life expectancies (DLE). Since this number cannot be observed until all insured in a portfolio have eventually died, a natural alternative is the difference in temporary life expectancies (DTLE). This number has the desired properties for old portfolios but is rather hard to interpret for young portfolios. We therefore additionally propose the normalized difference in temporary life expectancies

(NDTLE), and the indirect difference in life expectancies (IDLE). While NDTLE calculates an estimate for the ratio of the eventual deviation and an absolute or relative benchmark deviation that has to be set by the observer, IDLE is an intuitive estimate based on today's data for the DLE that will eventually materialize and is probably the easiest to interpret of all proposed metrics and has the most explanatory power. We apply our methodology to the database of an LE-provider and find that this underwriter's current methodology appears to be of high quality although he may have been slightly conservative in some recent years and more aggressive in very early years.

Our metrics should be of interest to all LE-providers when assessing their own quality of underwriting in order to identify weaknesses and improve their methodology. Also, investors in life settlements should use our metrics when assessing the quality of LE-providers, since the explained bias in A/E-ratios for old portfolios might conceal differences between LE-providers and might make LE-providers look better than they are. Moreover, when assessing the performance of their own portfolios, our metrics might be of use to investors. This is of particular interest when calculating the net asset value and the expected future return of a portfolio or fund: Given today's information, IDLE is a natural and intuitive estimate for the DLE that will eventually materialize in the portfolio when all insureds have died. Stressing all original LE-estimates by the value of IDLE and calculating the net asset value, the expected future return (or a return distribution) results in values that "automatically" correct for deficiencies in the original LE-estimates. A possible caveat to this statement is the implicit assumption that imperfections in the estimates are "homogeneous" in the sense that they only result from wrong frailty factors but there are no systematic deviations between the shapes of the mortality curves (this is also assumed in our illustrations in Sections 2 and 3). If we allow for arbitrary deviations between the "correct" mortality curve and the curve used in underwriting, it will be impossible to make an inference of the quality of eventual LEs and it will be impossible to assess underwriting quality with a single number. However, when positing that past underwriting performance is a guide to future underwriting performance, our metrics and particularly the IDLE overcome deficiencies of the industry standard and provide an accurate picture of the underwriter's quality.

References

Actuarial Standards Board (2013), "Life Settlements Mortality." Available at http://www.actuarialstandardsboard.org/wp-content/uploads/2014/02/asop048_175.pdf.

Aspinwall, Jim, Geoff Chaplin, and Mark Venn (2009). "Life Settlements and Longevity Structures: Pricing and Risk" New York: John Wiley and Sons.

Bauer, Daniel, Jochen Russ, and Nan Zhu (2016) "Adverse Selection in Secondary Insurance Markets: Evidence from the Life Settlement Market." Working Paper, Georgia State University.

Behan, Donald F. (2012) "Measurement of Incurred but Unreported Deaths in Life Settlements." ARC 2013.1 Proceedings.

Billingsley, Patrick (1995) "Probability and Measure." 3rd Edition. New York: John Wiley and Sons.

Bhuyan, Vishaal B. (2009) "Life Markets: Trading Mortality and Longevity Risk with Life Settlements and Linked Securities." New York: John Wiley and Sons.

Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt (1997) "Actuarial Mathematics." Second Edition, Schaumburg, Il: Society of Actuaries.

Braun, Alexander, Nadine Gatzert, and Hato Schmeiser (2012) "Performance and Risks of Open-End Life Settlement Funds." *Journal of Risk and Insurance*, 79: 193-230.

Davó, Nuria Bajo, Carmen Mendoza Resco, and Manuel Monjas Barroso (2013). "Portfolio Diversification with Life Settlements: An Empirical Analysis Applied to Mutual Funds." *The Geneva Papers on Risk and Insurance—Issues and Practice,* 38: 22-42.

Gatzert, Nadine (2010) "The Secondary Market for Life Insurance in the U.K., Germany, and the U.S.: Comparison and Overview." *Risk Management and Insurance Review*, 13: 279-301.

Giaccotto, Carmelo, Joseph Golec, and Bryan P. Schmutz (2015) "Measuring the Performance of the Secondary Market for Life Insurance Policies." *Journal of Risk and Insurance*, forthcoming.

Januário, Afonso V., and Narayan Y. Naik (2014). "Testing for adverse selection in life settlements: The secondary market for life insurance policies." *Available at SSRN.*

Qureshi, A. Hasan and Michael V. Fasano (2010) "Measuring Actual to Expected Accuracy for Life Settlement Underwriting." *Reinsurance News*, 68: 23-26.

Rhodes, Thomas E. and Stephen A. Freitas (2004) "Advanced Statistical Analysis of Mortality." Proceedings of the 14th International AFIR Colloquium.

Ruß, Jochen (2005) "Wie gut sind die Lebenserwartungsgutachten bei US-Policenfonds?" In: Florian Schoeller und Martin Witt (Editors): SCOPE Jahrbuch Geschlossene Fonds, 373-379.

Seligman, Edward J., and Sheldon Kahn (1980) "Testing for Significant Differences Between Actual and Expected Results." *Transactions of Society of Actuaries*, 32: 585-600.

Zhu, Nan, and Daniel Bauer (2013) "Coherent Pricing of Life Settlements Under Asymmetric Information." *Journal of Risk and Insurance*, 80: 827-851.

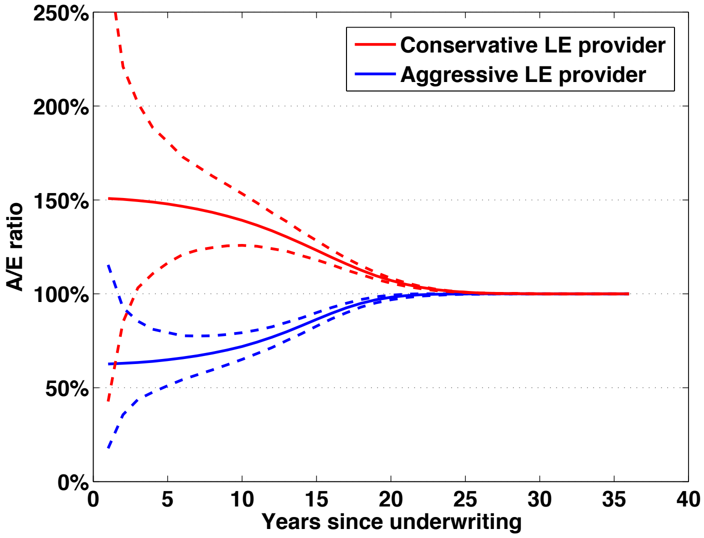

Figure 1: The development of the A/E ratios for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N = 500.*

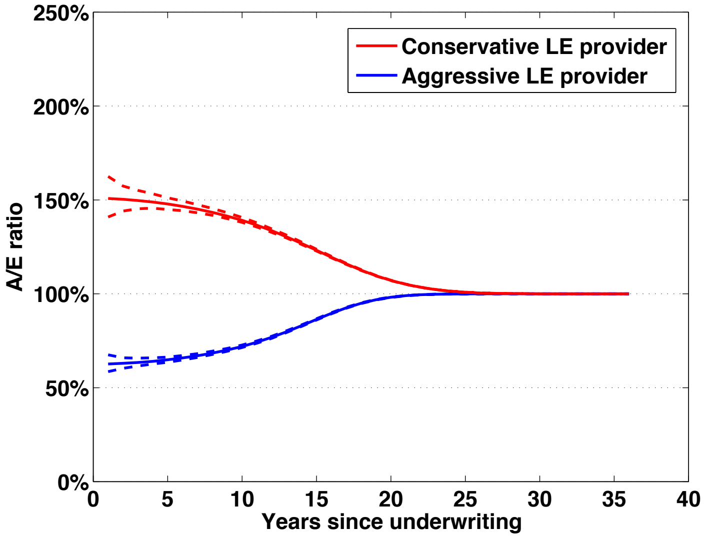

Figure 1: The development of the A/E ratios for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N = 50,000.*

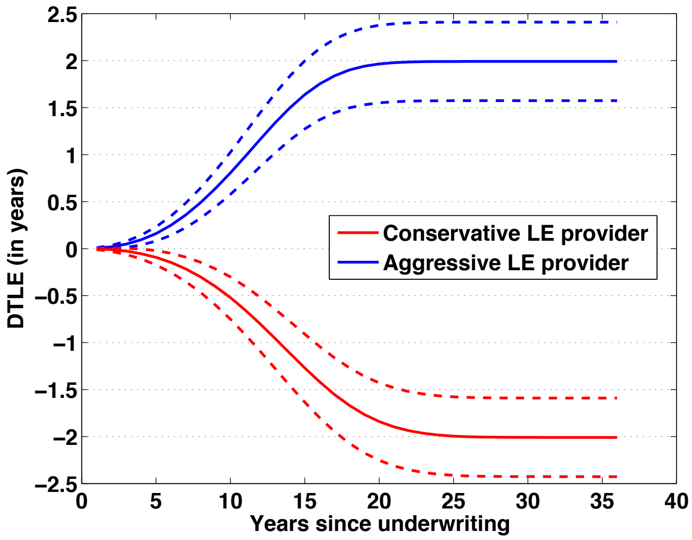

Figure 3: Development of DTLE for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N=500*.

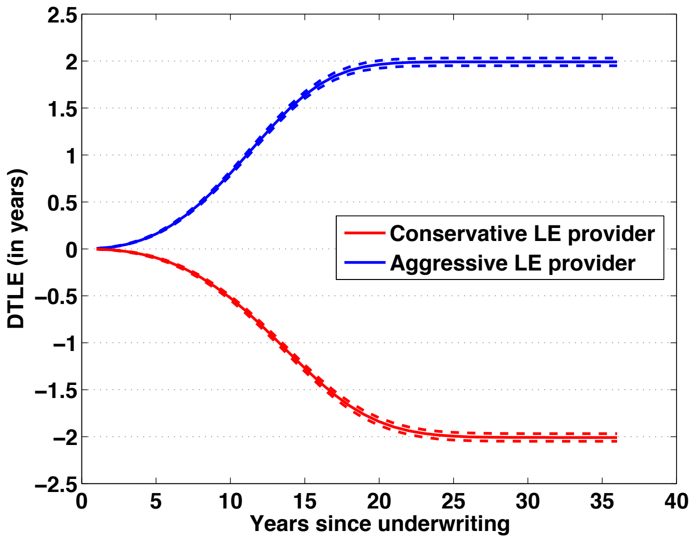

Figure 4: Development of DTLE for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N=50,000*.

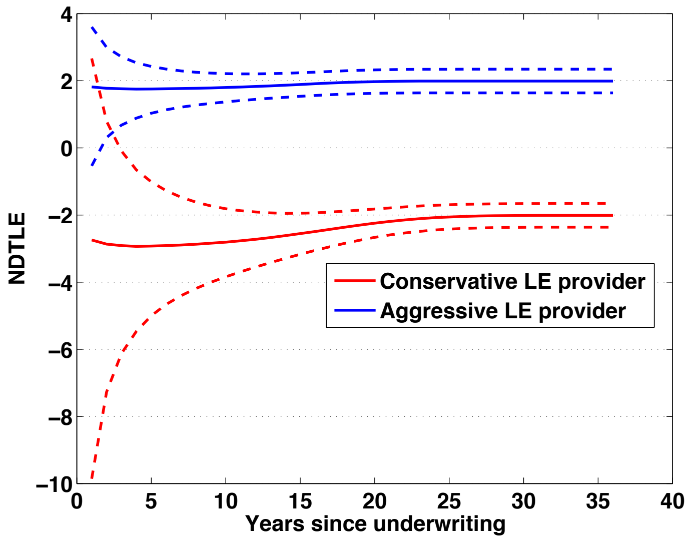

Figure 5: Development of NDTLE$_{d=1\ year}$ for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N=500.*

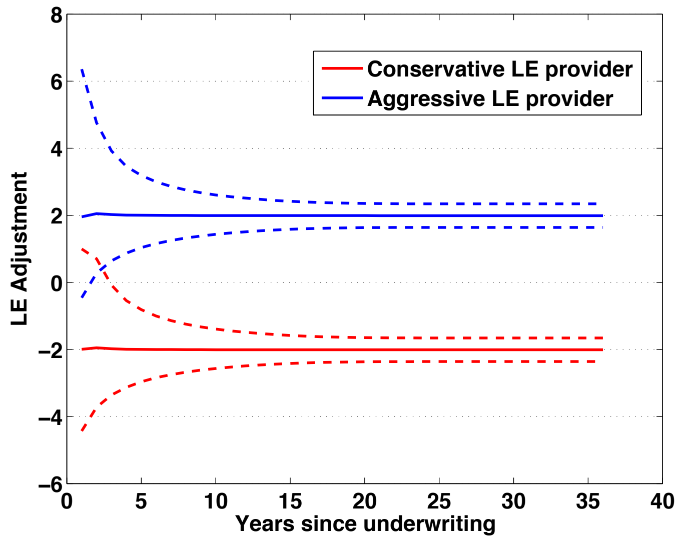

Figure 6: Development of IDLE for the aggressive LE provider (A) [blue solid line] with associated 2.5% and 97.5% percentiles [blue dash lines] and the conservative LE provider (C) [red solid line] with associated 2.5% and 97.5% percentiles [red dash lines] in the hypothetical example. *N=500.*

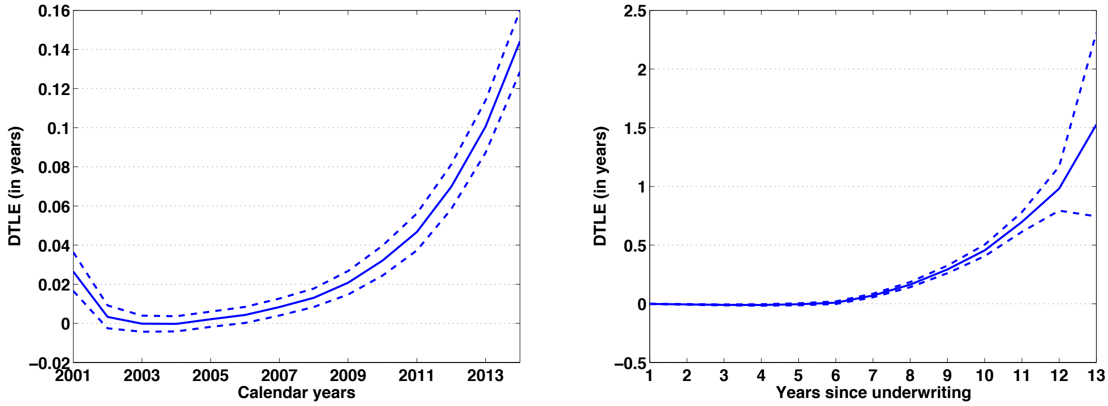

Figure 7: The development of DTLE for the Fasano dataset as a function of calendar year (left panel) or year since underwriting (right panel) (blue solid line) with associated 2.5% and 97.5% percentiles (blue dash lines).

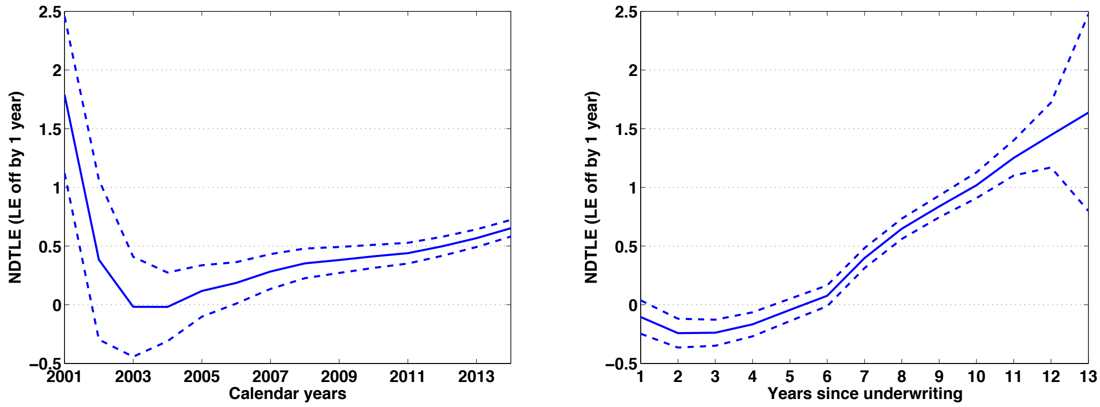

Figure 8: The development of $NDTLE_{d=1\ year}$ for the Fasano dataset as a function of calendar year (left panel) or year since underwriting (right panel) (blue solid line) with associated 2.5% and 97.5% percentiles (blue dash lines).

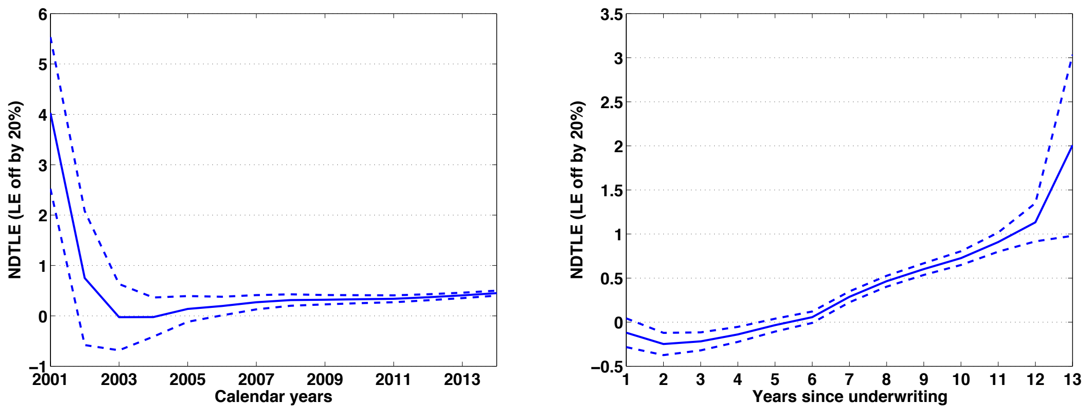

Figure 9: The development of $NDTLE_{p=20\%}$ for the Fasano dataset as a function of calendar year (left panel) or year since underwriting (right panel) (blue solid line) with associated 2.5% and 97.5% percentiles (blue dash lines).

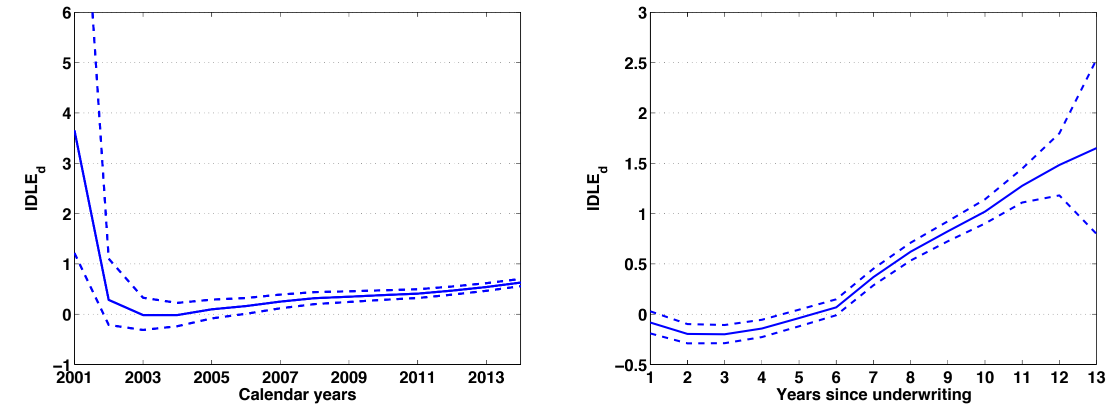

Figure 10: The development of IDLE$_d$ for the Fasano dataset as a function of calendar year (left panel) or year since underwriting (right panel) (blue solid line) with associated 2.5% and 97.5% percentiles (blue dash lines).

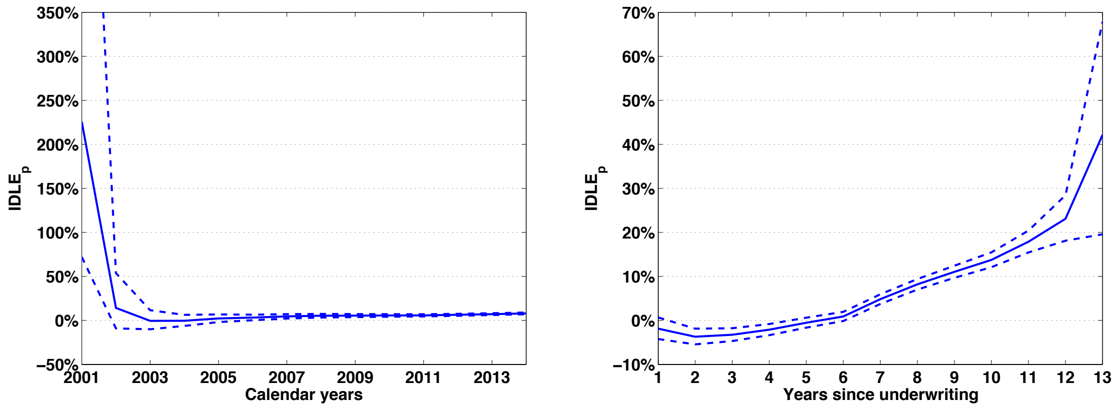

Figure 11: The development of IDLE$_p$ for the Fasano dataset as a function of calendar year (left panel) or year since underwriting (right panel) (blue solid line) with associated 2.5% and 97.5% percentiles (blue dash lines).

	Entire Portfolio	Before January 1st 2006	After January 1st 2006
DTLE	0.1442	0.4233	-0.0049
95% C.I.	(0.1286, 0.1598)	(0.3863, 0.4603)	(-0.0188, 0.0091)
NDTLE$_{d=1}$	0.6521	1.0281	-0.0385
95% C.I	(0.5815, 0.7227)	(0.9382, 1.1180)	(-0.1487, 0.0717)
NDTLE$_{p=20\%}$	0.4512	0.7163	-0.0263
95% C.I	(0.4023, 0.5001)	(0.6537, 0.7789)	(-0.1017, 0.0491)
IDLE$_d$	0.6283	1.0308	-0.0342
95% C.I	(0.5562, 0.7015)	(0.9328, 1.1306)	(-0.1306, 0.0639)
IDLE$_p$	0.0798	0.1352	-0.0041
95% C.I	(0.0705, 0.0894)	(0.1219, 0.1488)	(-0.0156, 0.0077)

Table 1: Summary statistics for DTLE, NDTLE$_{d=1}$, NDTLE$_{p=20\%}$, IDLE$_d$, and IDLE$_p$ and their associated 95% confidence intervals. The metrics are calculated as of January 1st 2015 for the entire Fasano portfolio, as well as two sub-portfolios that were underwritten before and after January 1st 2006, respectively.