Stockholms universitet

Evaluation of Model Assisted Survey Sampling for Life Insurance Technical Provision Calculations

Mattias Lundahl

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2017:5
Försäkringsmatematik
Maj 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2017:5**
http://www.math.su.se

Evaluation of Model Assisted Survey Sampling for Life Insurance Technical Provision Calculations

Mattias Lundahl*

May 2017

Abstract

This thesis aims to evaluate the feasibility of using model assisted survey sampling to estimate the technical provisions of a traditional life insurance portfolio, for IFRS purposes as well as Solvency II best estimate purposes. The goal is to achieve an estimate of technical provisions of certain subsets of the portfolio that falls within a certain tolerance margin of the true provisions calculated using the entire portfolio, while using only a limited subset of policies to significantly shorten the calculation's running time. The evaluation compares two separate survey sampling techniques, estimating the aggregate technical provisions in the subsets of the portfolio firstly by scaling individual policies' calculated technical provisions based on their respective selection probabilities and secondly by fitting a multiple linear regression to the selected policies (where the surrender value and second order reserve predict the technical provisons) and use this model to explicitly predict the technical provisions of each individual policy not included in the sample. In particular the regression based estimator is found to be accurate to within less than 0.5% of the true aggregate technical provisions of each subset on average, even for sample sizes as low as 3% of the total portfolio in terms of number of policies. Since the calculation time for technical provisions is approximately linear in the number of policies used by the provision model, the suggested sampling method can save significant amounts of time. The regression model variables and selection probability parameter values are found to be robust when tested on other time periods, indicating that frequent recalibration would not be required. Using unequal selection probabilities based on each policy's surrender value provides an added benefit to the trade off between accuracy and calculation time compared to using a fixed probability for each policy.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: mattias.lundahl@gmail.com. Supervisor: Filip Lindskog.

Acknowledgements

I would like to thank my supervisor Filip Lindskog of Stockholm University for valuable input regarding the underlying theory and suggestions of useful references. I would also like to thank those of my colleagues who have provided feedback on draft versions of this thesis and thoughts on reasonable choices of parameters for the sample selection model.

Contents

Chapter 1

Introduction

According to Swedish law and the Solvency II directive, insurance companies need to determine their technical provisions on a recurring basis to ensure sufficient capital is held to cover future payment liabilities. The calculation of technical provisions can be a time consuming task for a life insurance portfolio, where the amount paid to policy holders depends on factors such as whether the policy holder has kept up premium payments, moved the policy, died (before or after the intended start of payouts) or whether the policy has reached its maturity, as well as policy specific details such as different technical bases and taxation classes.

To calculate technical provisions for a life insurance portfolio requires establishing a state model for how, when and with what probability policies can transition between states. A model for future cash flows into and out of an insurance company, in each of the states that policies can belong to, is also necessary. Finally, a model for the time value of money enables determining the present value of future liabilities.

The portfolio of primary interest for the purposes of this thesis is the traditional life insurance portfolio of a particular large Swedish insurance company. This portfolio contains a wide range of different policy tariffs, technical bases, mortality bases, policy holder ages and various other conditions. For this reason, the process of calculating provisions is lengthy. To ensure that financial reports can be prepared in a timely fashion amid increased regulatory demands, methods which can reduce computational complexity while maintaining sufficient accuracy are in high demand.

To keep the evaluation focused on sample selection techniques rather than the minutiae of numerous less popular policy tariffs, this thesis will consider a subset of more common types, specifically annuities with a fixed maximum duration and annuities over the life of the policy holder, with and without a beneficiary in the event that the policy holder dies. Joint life policies will not be included in the analysis, because they are much less common in the portfolio and also require a more complicated state model.

The provision calculation model will be described in sufficient detail to demonstrate the cause of long calculation times, together with an overview of the insurance mathematics included in the model. The background of the method used to select a subset of representative policies will also be detailed.

The rest of the thesis is organized as follows: Chapter 2 gives a brief review of basic life insurance mathematics and introduces the provision calculation model. In Chapter 3, the foundations of model assisted survey sampling are discussed. The application of the method to this particular case is covered in Chapter 4 with details of the implementation. Results and conclusions are presented in Chapter 5. Finally, appendix A provides a more in-depth, illustrated explanation of the various components of the provision model and how these work together, while appendix B provides an overview of the results of the parameter tests conducted for the sampling method.

Chapter 2

Mathematical background

2.1 A general introduction to technical provisions

Swedish law [3] for the insurance industry defines the obligation of companies in this industry to periodically calculate their technical provisions according to certain legal standards and the company's own technical reference documents. At the European level, the Solvency II directive [2] specifies essentially similar requirements. The purpose of calculating technical provisions is to ensure that the insurance company has enough capital set aside to cover future payment liabilities. Such liabilites exist in the form of benefit payments to policy holders and beneficiaries, operating expenses and taxes.

For policies where the benefits are calculated using an assumption of recurring premiums, the expected future premiums themselves reduce the required technical provisions, because these expected cash flows increase the capital held by the insurance company to use for payments to the particular policy holder or beneficiary. In the event that premium payments were to cease for such policies, the benefits will also be lowered accordingly. Hence, subtracting expected future premiums from the capital requirements does not in and of itself create a risk of insufficient capital in the future.

Each of these four cash flow components can hypothetically be modeled as stochastic processes. If cash flows occur at discrete points in time, the benefit claims can be denoted $C(t)$, the operating expenses $E(t)$, the taxes $X(t)$ and the future premiums $P(t)$, where each of the four are stochastic variables that can take different non-negative values at each time t.

Calculating technical provisions then becomes a task of determining a net present value of the expected values of these cashflow components for all future points in time. Denoting the net cash flow at each point in time by $N(t)$, we have $N(t) = C(t) - P(t) + E(t) + X(t)$ and with $PV_t(x)$ denoting the present value at tine t of a future cash flow x, the technical provisions at time t' are

$$tp(t') = \sum_{t=t'}^{\infty} E\left[PV_{t'}(N(t))\right]$$

The most relevant technical provisions are generally those at the current point in time,

$$tp(0) = \sum_{t=0}^{\infty} E\left[PV_0(N(t))\right]$$

This notation allows for the discounting used to determine the present value to itself have a stochastic quality, i.e. to use a stochastic rate model or even model dependence between the cash flow quantities and the discount rates. In practice, the rates might be considered deterministic, in which case the present value "operator" could equivalently be applied to calculated expected future cash flow values.

A fully general model for calculating technical provisions might make assumptions about a probability distribution of each cash flow component at each point in time and use known expressions for the expected value or a Monte Carlo simulation based approach if appropriate.

It would also be possible to model the expected values of cash flows directly, without explicit assumptions about specific distributions that the cash flows might follow. This can for instance be useful for so called *traditional life insurance* (see e.g. section 5.8 of Møller and Steffensen [5]), where the guaranteed part of benefit payments (the part that impacts technical provisions) is a known amount for each possible state and the only uncertainty is what state the policy will be in at each point in time. In this case, the expected values of the cash flows can be calculated by using the known amounts in each state together with the probabilities of being in each state. For notational simplicity, the expected values can be directly denoted by $c(t) = E[C(t)]$, $p(t) = E[P(t)]$, $e(t) = E[E(t)]$ and $x(t) = E[X(t)]$.

In either case, calculating the expected cash flow components at any particular point in time will require an idea of what state each particular policy holder is in, since benefit claims, expenses, taxes and premiums will typically depend on whether the policy holder is alive or dead, has transferred the policy out of the insurance company, has ceased or continued premium payments or whether the policy has started paying out or reached its maturity (if any).

Keeping track of the states requires some form of state model, which could be implemented in the form of a Monte Carlo simulation with multiple runs where the policy holder is in one particular state at each point in time for each run. Alternatively, the state model might be implemented in the form of a Markov chain, where a probability of being in each state at each point in time is determined and the (expected) cash flow components are essentially calculated for each state and each time t and then aggregated into an overall expected value for each time t.

In the most general case, a continuous time Markov chain could be used, where the probabilities that the policy holder is in each of the various states at time t can be denoted by the vector

$$\mathbf{p_S}(t) = \begin{bmatrix} p_1(t) & p_2(t) & \dots & p_N(t) \end{bmatrix},$$

where N is the total number of different states.

It might also be possible to simplify this somewhat to a discrete time Markov chain, in case policy events are only actually handled on a periodic basis. It is common in practice that life insurance companies use monthly periodization, where interest is added and expenses are subtracted once per month and benefits are paid monthly or less frequently, while premiums can in principle be paid more frequently.

In this case, the same notation can be used to denote the probabilities that the policy holder is in each of the given states at time t (which is in this case an integer valued variable). This notation is similar to that of Section 4.4 in Ross [7], though in this case the probabilites are stated at a given point in time rather than stating the long run probabilities of being in each state. Additionally, in this case it is relevant to introduce a transition matrix $P_S(t)$ according to

$$P_S(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \dots & p_{1N}(t) \\ p_{21}(t) & p_{22}(t) & \dots & p_{2N}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1}(t) & p_{N2}(t) & \dots & p_{NN}(t) \end{bmatrix},$$

where each $p_{ij}(t)$ denotes the probability that the policy holder will transition from state i to state j from time t to $t+1$. This means that the following relationship holds between the state probabilities at times t and $t+1$ and the transition matrix:

$$\mathbf{p_S}(t+1) = \mathbf{p_S}(t) P_S(t),$$

where the usual Markov chain convention of using a row vector for the state probabilities is used.

It is also possible to reformulate this multiplicative, matrix based notation into an equivalent additive notation for each state separately. For details on how to derive this reformulation, see section A.2.

2.2 Overview of basic life insurance mathematics

Fundamentally, calculating the technical provisions for life insurance policies requires the ability to value individual payments and annnuity payments while simultaneously taking both mortality and the time value of money into account. As described in e.g. Andersson [1], this requires defining the survival function,

$$l(x) = P(\text{Policy holder still alive at duration x}),$$

from which the single time period mortality probability can be shown to be

$$q_x = 1 - \frac{l(x+1)}{l(x)}$$

Literature on the topic of life insurance mathematics will frequently consider this to be the probability of death within one year, although it is equally possible to use a different time unit. The model investigated in this thesis uses a time unit of one month.

To value future single payments or annuity payments, it is customary to introduce commutation factors to encapsulate important recurring calculations in shorthand notation. The so called "discounted number of the living" is used to calculate the value of a future single payment conditional on the recipient remaining alive at the time of payment. This is defined as

$$D(x) = l(x)e^{-\delta x}$$

To calculate the value of an annuity payment conditional on the recipient remaining alive, the so called "sum of the discounted number of the living" is introduced according to

$$N(x) = \int_x^\infty D(t)dt$$

This number by itself actually relates to a hypothetical perpetuity, as indicated by the upper integration limit. To determine the value of an annuity with a maximum length, differences of two such numbers can be used.

The Makeham model is used to model mortality, meaning the intensity of mortality has the form

$$\mu_x = \alpha + \beta e^{\gamma(x-f)}, x \geq 0.$$

The relationship between the mortality intensity μ_x and the survival function $l(x)$ is

$$\mu_x = -\frac{l'(x)}{l(x)} \Leftrightarrow l(x) = e^{-\int_0^x \mu_s ds}$$

2.3 Brief description of the provision model

The core of the provision model studied in this thesis is the state model, a discrete Markov chain where the policy holder can belong to either of the following states:

Short form name	Description of state
PP	Premium paying
PU	Paid-up
POPP	Paying out, from premium paying
POPU	Paying out, from paid-up
TR	Transferred
MA	Matured
DE	Dead

Table 2.1: Possible states for the state model

The reason for keeping track of the states POPP and POPU separately rather than simply a single state PO is that policies with an explicit condition of recurring premium payments will receive a reduced benefit in case the policy holder decides to cease premium payments - the original estimated benefit for these policies is calculated based on the assumption that premium payments will be made according to the condition at the issue date. Hence, the calculated benefit present in the raw policies data used in the model (see section A.5) will not be accurate for such policies and will need to be recalculated. By keeping POPP and POPU separate states, the benefit can be determined for each of these separately.

The possible transitions between these states are described in the following figure (note though that transitions directly from PP or PU to MA are only possible for policies with a single payout at the time of maturity):

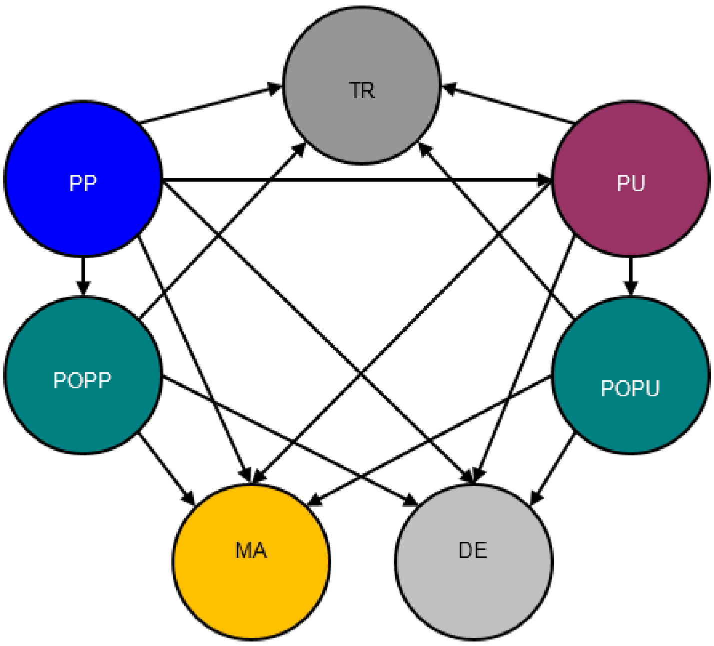

Figure 2.1: Possible states and transitions in the state model

In order to calculate provisions, the model attempts to calculate the function reflecting the technical provisions at time 0. This depends on the cash flow components benefit claims, premiums, operational expenses and taxes, as well as the technical provisions themselves, all at time 1. The result will be discounted back one time period.

Since the insurance portfolio investigated consists of traditional life insurance, policy holders will receive guaranteed payments and possibly discretionary bonus amounts depending on the performance of the asset portfolio. Only the guaranteed payments affect the technical provisions, because the bonus amounts are discretionary and can be withheld if the asset portfolio has underperformed and the aggregate capital is insufficient.

The guaranteed payments are calculated by the source system at issuance (for policies with an explicit condition of recurring premiums or policies with a purely one-off premium) or recalculated as each premium is paid (for policies that belong to neither of the aforementioned categories), and the calculated amount is provided as data to the provision model (see section A.5). For policies in the last category, no future premiums are assumed to occur in the provision model. This means that the cash flow components can be treated as deterministic, independently of asset performance as well as discount rates.

Because of the recursive dependence on itself (a priori, the technical provisions are only known to be 0 at the end of the projection period where no future cash flows exist) and because the other constituent variables in turn depend on the state model (which inherently needs to be calculated forward in time before the expected cash flows), the computational complexity of the model is considerable. A more in-depth, illustrated elaboration of the complexity is provided in appendix A, sections A.1 to A.3.

Furthermore, the model calculates both the provisions of each individual policy at time 0 as well as the aggregate cash flow components (claims, premiums, expenses and taxes) of different groups of policies for each of up to 1 440 months (the model assumes that the maximum age any policy holder will reach is 120 years). This is because certain regulatory reports require knowledge of the future cash flow profile in various parts of the portfolio - for instance occupational pension and other life insurance separately, P and K tax classes separately etc (see section A.3). To have this data available requires calculating each cash flow component at each of 1 440 points in time for each individual policy.

In conclusion, the model needs to project each individual policy forward in time up to 120 years into the future using a monthly discrete Markov chain, calculate the payout amount in the different states at each point in time and discount all these payments back to the starting point using the relevant interest rate curve. With multiple different states and a total portfolio of roughly one million policies, the total calculation time is currently about an hour for the base case.

Considering regulatory requirements for multiple different scenarios (see for instance the regulation regarding traffic light reporting from Finansinspektionen [9] or section SCR.7 of the QIS5 Technical Specifications [6] for Solvency II), adjusting the rates of mortality and/or other parameters, the total time spent calculating provisions in these scenarios quickly increases.

2.4 Description of the policy tariff types

The focus of the analysis will be the very popular annuity policies, both life-long and with a fixed maximum duration and with/without a beneficiary to receive payments in the event that the policy holder dies before the policy has matured. These types correspond to the "term/permanent life insurance" referred to in section 1.2.1 of Koller [4] and are commonly known by the following short form tariff names:

Short form name	Tariff type
R	Life-long annuity, no beneficiary
TR	Fixed maximum duration annuity, no beneficiary
RSH	Life-long annuity, fixed term annuity to beneficiary
TRSH	Fixed maximum duration annuity, same duration for beneficiary

Table 2.2: Tariff types considered in the analysis

The R tariff corresponds to section 3.4.3 of Andersson [1], meaning that the value of future benefits can be expressed as

$$A(t) = \begin{cases} L \cdot \frac{N(x+k)}{D(x+t)} & 0 < t \leq k \\ L \cdot \frac{N(x+t)}{D(x+t)} & k < t, \end{cases}$$

where k is the remaining time until benefit payments start, and the value of future premiums (for policies with PRM_FLAG = L only, see section A.5) is

$$B(t) = \begin{cases} P \cdot \frac{N(x+t)-N(x+n)}{D(x+t)} & 0 < t \leq n \\ 0 & n < t \leq k \\ 0 & k < t, \end{cases}$$

where n is the remaining time during which premiums are assumed to be paid.

The TR tariff similarly corresponds to section 3.4.4, meaning that the value of future benefits is

$$A(t) = \begin{cases} L \cdot \frac{N(x+k)-N(x+k+s)}{D(x+t)} & 0 < t \leq k \\ L \cdot \frac{N(x+t)-N(x+k+s)}{D(x+t)} & k < t \leq k+s \\ 0 & k+s < t, \end{cases}$$

where s is the maximum duration of the period during which benefits will be paid, and the value of future premiums (for policies with PRM_FLAG = L only, see section A.5) is

$$B(t) = \begin{cases} P \cdot \frac{N(x+t)-N(x+n)}{D(x+t)} & 0 < t \leq n \\ 0 & n < t \leq k \\ 0 & k < t. \end{cases}$$

The RSH and TRSH tariffs are variants of R and TR respectively, where benefits will be paid to a beneficiary in case the policy holder dies, either before benefits are scheduled to start paying out or within a certain amount of time after benefit payments have started. The valuation of these tariffs *does not* require the beneficiary to be alive, meaning these are *not* joint-life tariffs. In the event that a beneficiary is entitled to benefit payments, these are considered unconditional cash flows in the calculation of technical provisions.

The derivations of the above expressions in Andersson [1] assume a constant and deterministic interest intensity δ. In practice, the formulas for calculating the values of claims and premiums presented above are modified by introducing discretization adjustments as well as administrative loadings, and present values of future cash flows are calculated using deterministic interest rate curves rather than a constant rate. Only the surrender value, appearing in the provision model specifically for the case where the policy holder transfers the policy out of the insurance company, and the guaranteed payment amount generated by a premium payment, are calculated assuming a constant interest rate intensity using formulas similar to those above. For further details, see section A.1.

Chapter 3

Model assisted survey sampling

The underlying idea of model assisted survey sampling is to let features of a model and of its data influence the decision of how to sample data to obtain a representative subset of data points that can be used to make inferences about the population from which samples are drawn.

As described by Särndal, Swensson, Wretman [8], there are multiple different approaches to selecting samples from a population. Fundamentally, the *population* (denoted by $U = \{1, \dots, k, \dots, N\}$) is the full set of data points for which inferences are to be made. When conducting a sample selection, a *sampling frame* is any material or device used to enable observations in the population at hand. An individual *sample* is denoted by S.

A fundamental property of sampling is the probability of inclusion for a given element k from the sampling frame. By introducing an indicator variable

$$I_k = \begin{cases} 1 & \text{if } k \in S \\ 0 & \text{otherwise,} \end{cases}$$

the inclusion probability π_k for element k can be expressed as

$$\pi_k = P(k \in S) = P(I_k = 1).$$

Frequently, it is relevant to find estimates of the aggregate value of some element specific quantitative variable y_k. This aggregate can be denoted by $t = \sum_{k \in U} y_k$ and estimated using common statistical techniques based on model assumptions and sampling method. In general, the π estimator is defined using the inclusion probability as

$$\hat{t}_\pi = \sum_{k \in S} \frac{y_k}{\pi_k}.$$

As shown by Särndal et al [8], this estimator is unbiased for t with variance given by

$$Var(\hat{t}_\pi) = \sum_{k \in U} \sum_{l \in U} \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l},$$

where Δ_{kl} is the covariance between I_k and I_l, i.e.

$$\Delta_{kl} = Cov(I_k, I_l) = E[I_k I_l] - E[I_k]E[I_l] = \pi_{kl} - \pi_k \pi_l.$$

3.1 Approaches to sample selection

In principle, the population might be more or less clearly defined, introducing the need for different sampling methods in the various scenarios. The most straight forward case is when the population is fully available in compiled format (such as a computer file), in which case the population and the sampling frame can be considered identical. In other cases, a sampling frame will need to be designed separately.

With a sampling frame established, the sampling method will further depend on whether there are subgroups within the population and whether such subgroups have a natural hierarchical organizational model. For instance, a survey conducted in a country should cover respondents from various regions of the country and possibly also from various cities within regions, requiring a *cluster-based* approach to sampling. In the absence of hierarchical patterns, it is still frequently relevant to *stratify* the population to cover subsets which may have stratum specific properties.

On a more basic level, samplings within the (possibly clustered or stratified) population can be conducted through various methods, each suitable for a different scenario. Below are brief descriptions of some of these approaches to sampling.

3.1.1 Bernoulli sampling

This sampling method means that each element of the sampling frame has a fixed probability π of being included in the sample, and hence a fixed probability $(1-\pi)$ of not being included. Clearly, the sample size in this case follows a $Bin(N, \pi)$ distribution. In this case, the π estimator of the population total $t = \sum_{k \in U} y_k$ will take the form

$$\hat{t}_\pi = \frac{1}{\pi} \sum_{k \in S} y_k.$$

3.1.2 Simple random sampling

If it is more relevant to ensure a uniform sample size and if each sample of this size should have equal probability, simple random sampling without replacement is an appropriate method. For a desired sample size n, the probability of inclusion for each element is n/N and the estimator of the population total becomes

$$\hat{t}_\pi = \frac{N}{n} \sum_{k \in S} y_k.$$

Simple random sampling with replacement will instead assign equal probability $1/N^m$ to each ordered sample of size m.

3.1.3 Systematic sampling

For simplicity of execution, systematic sampling is an appropriate choice. The only randomization required is to pick a first element among the a first elements in the sampling frame, after which each ath element is added to the sample. If the population size N is an integer multiple of a, samples will have a common size. Otherwise, sample sizes may differ by one element. The estimator of the population total is

$$\hat{t}_\pi = a \sum_{k \in S} y_k.$$

3.1.4 Poisson sampling

In case equal probability sampling is not desirable, Poisson sampling is a useful and simple approach. Setting the inclusion probability of element k as π_k, the sample size is random with expected value $\sum_{k \in U} \pi_k$ and the estimator of the population total is the general one introduced earlier, i.e.

$$\hat{t}_\pi = \sum_{k \in S} \frac{y_k}{\pi_k}.$$

3.1.5 Probability proportional-to-size sampling

This is a particular implementation of Poisson sampling using a probability proportional to the size of some auxiliary variable x_k that is itself approximately proportional to the objective variable y_k. Since this is a special case of Poisson sampling, the same expression holds for the estimator of the population total.

3.1.6 Stratified sampling

The technique of stratification requires identifying relevant, disjoint strata based on quantitative or categorical data relating to the elements in the population. Each stratum requires a sample size and sampling design, which can be uniform or stratum specific. With a total of H strata, the estimator of the population total becomes

$$\hat{t}_\pi = \sum_{h=1}^{H} \hat{t}_{h\pi},$$

where $\hat{t}_{h\pi}$ refers to the corresponding estimator for stratum h.

3.2 Auxiliary variables and the regression estimator

The estimation accuracy for the aggregate value of the y_k variable can be further improved through the use of multiple *auxiliary variables* in a *regression estimator*. This requires identifying J auxiliary variables $\mathbf{x}_k = (x_{1k}, \ldots, x_{Jk})^T$ with sufficient predictive power for the objective variable y_k and use known values of these auxiliary variables for the elements *not* included in a particular sample in order to determine specific estimates of the y_k values of these elements.

The general regression estimator is denoted by $\hat{t}_{yr}$ and defined as

$$\hat{t}_{yr} = \hat{t}_{y\pi} + \sum_{j=1}^{J} \hat{B}_j \left(t_{x_j} - \hat{t}_{x_j\pi} \right),$$

where

$$\hat{t}_{y\pi} = \sum_{k \in S} \frac{y_k}{\pi_k}$$

is the usual π estimator of $t_y = \sum_{k \in U} y_k$,

$$\hat{t}_{x_j\pi} = \sum_{k \in S} \frac{x_{jk}}{\pi_k}$$

is the π estimator of the known x_j total

$$t_{x_j} = \sum_{k \in U} x_{jk}$$

and $\hat{B}_1, \ldots, \hat{B}_J$ are components of the vector

$$\hat{\mathbf{B}} = \left(\sum_{k \in S} \frac{\mathbf{x}_k \mathbf{x}_k^T}{\sigma_k^2 \pi_k} \right)^{-1} \sum_{k \in S} \frac{\mathbf{x}_k y_k}{\sigma_k^2 \pi_k}.$$

An alternate expression for the regression estimator can be obtained by noting that the fitted values $\hat{y}_k$ of the objective variable can be written

$$\hat{y}_k = \mathbf{x}_k^T \hat{\mathbf{B}} = \sum_{j=1}^{J} \hat{B}_j x_{jk}$$

and the sample fit residuals as

$$e_{kS} = y_k - \hat{y}_k.$$

The regression estimator becomes

$$\hat{t}_{yr} = \sum_{k \in U} \hat{y}_k + \sum_{k \in S} \frac{e_{kS}}{\pi_k}.$$

With the help of this alternate expression and some further notation, an approximate variance and a variance estimator for the regression estimator can be determined. Hence, confidence intervals can be formed for the population total through the ordinary normal approximation.

Chapter 4

Survey sampling for the provision model

For the particular provision model studied in this thesis, the full population of policies is available at the outset as a computer file, meaning that the sampling frame is identical with this file. The ultimate aim is to estimate an aggregate of the individual policy technical provisons, meaning that these technical provisions represent the y_k values whose sum is the object of interest. Apart from the total sum, partial sums over certain segments of the portfolio are also required for regulatory purposes.

4.1 Application of survey sampling

As reflected in section A.3, the provision model studied in this thesis relies on a number of categorical variables to determine which values to use for certain parameters in the provision calculations (discount rate curves, expense and tax loadings, assumptions for policy holder options, etc.). These categorical variables will clearly be useful in the random sampling of policies, also because aggregate technical provisions need to be calculated for portfolio subsets defined by these categorical variables (for instance, separate aggregates are required for collectively negotiated occupational pensions, individual occupational pensions and other life insurance, as well as the two tax classifications).

Considering that geographical locations of policy holders or other hierarchical divisions are not relevant for the calculation of technical provisions (meaning that clustered designs are not appropriate) and that total provisions need to be estimated for well defined, disjoint groups of policies, the stratified sampling method appears most appropriate.

In general, the technical provisions will depend on calculated benefit amounts and the total capital contained in the policy. Hence, unequal sampling probability would appear most appropriate, with higher inclusion probability assigned to policies with more capital. Furthermore, the regression estimator is likely to produce improved estimates of the technical provisions of individual policies that are not included in the sample and hence a more reliable estimate of total technical provisions.

Since all policy data is available when conducting technical provision calculations, the concept of nonresponse is not applicable in this case. Measuring error is also not quite applicable in the precise sense considered by Särndal et al [8], although manual inputs to policies' data by administrative staff inevitably create a risk of introducing errors. Certain known issues related to delays in the handling of policy events (e.g. deaths of policy holders) are dealt with separately outside of the core model. These corrections are considered out of scope for this thesis.

4.2 Detailed implementation

A natural way to determine policy inclusion probabilities π_k would be to use a probability that increases with the size of benefit payments. An easy and intuitively reasonable way to implement this would be to use the variable $SV1ST$ from the policy data (see section A.5), denoting the surrender value paid out in case the policy is surrendered or transferred. Since the goal is to determine probabilities, the *logistic function*, which generally takes the form

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}},$$

seems a reasonable choice of transform. To obtain probabilities, L should have the value 1. Since $SV1ST$ is non-negative, x_0 (the parameter determining the horizontal point of symmetry, where the inclusion probability is 0.5) should have a sufficiently high positive value that a suitable number of policies are assigned inclusion probabilities of 0.5 or above. The steepness parameter k does not have an a priori obviously suitable value. Various combinations of values for x_0 and k will be evaluated to find a useful trade-off between calculation time and estimation accuracy.

The implementation of the regression estimator for policies that are not included in the sample can benefit from the fact that policy data includes more information than the parameters directly used by the provision model (the variables are listed in section A.5). Variables likely to be predictive of a policy's technical provisions include

- $SV1ST$, surrender value as calculated by source system
- RES_2ND, total capital (including discretionary bonus)
- RES_1ST, premium reserve
- BEN_1ST, benefit amount

4.3 Evaluation of logistic function parameters

Firstly, the subset of the portfolio consisting of the tariffs described in section 2.4 is analyzed to ascertain whether all of the accounting products (see section A.3 for details) are sufficiently represented to allow sampling with reasonable predictive capability. With the tariffs chosen, it turns out that accounting products B01, B03 and B15 are well represented while accounting products B02 and B04 have very few policies of either of these tariffs. The subsequent analysis will hence focus on accounting products B01, B03 and B15 (which are incidentally the accounting products with tax class P - see section A.3).

Since these accounting products may exhibit different properties, stratified sampling is conducted with the accounting products as strata. Each accounting product has a separate sample with potentially separate parameters for the logistic function as well as a separate regression analysis for the purposes of the regression estimator.

In order to determine suitable parameter values for the logistic function, all 100 combinations of the following values of k and x_0 are tested for a single sample per accounting product, using a common random seed:

k	x_0
10^{-1}	10000
10^{-2}	20000
10^{-3}	30000
10^{-4}	40000
10^{-5}	50000
10^{-6}	60000
10^{-7}	70000
10^{-8}	80000
10^{-9}	90000
10^{-10}	100000

Table 4.1: Tested values for logistic function parameters

In each case, the π estimator and the regression estimator are used to construct estimates of the accounting product aggregate technical provisions. For the regression estimator, it generally holds that combinations of the four parameters listed in section 4.2 above provide highly significant fits with adjusted R-square values of 0.96 or above as long as at least one of the first three is included. Each individual parameter also is highly significant with a p value of less than $2 \cdot 10^{-16}$, though the adjusted R-square does not change much by including more of these variables. By the principle of parsimonious models, the common model

$$y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + \epsilon_k$$

will be used, where y_k is the technical provisions of policy k and x_{1k} and x_{2k} correspond to the surrender value $SV1ST$ and second order provisions (including discretionary bonus) RES_2ND, as these are intuitively reasonable as predictors of a policy's technical provisions.

Running these tests for one sample and one regression estimation per accounting product and combination of k and x_0 produces relative errors (calculated as $\left|\frac{\hat{t}_\pi}{t} - 1\right|$ and $\left|\frac{\hat{t}_{yr}}{t} - 1\right|$ respectively) compared to the calculated true aggregated technical provisions as reflected in the following figure:

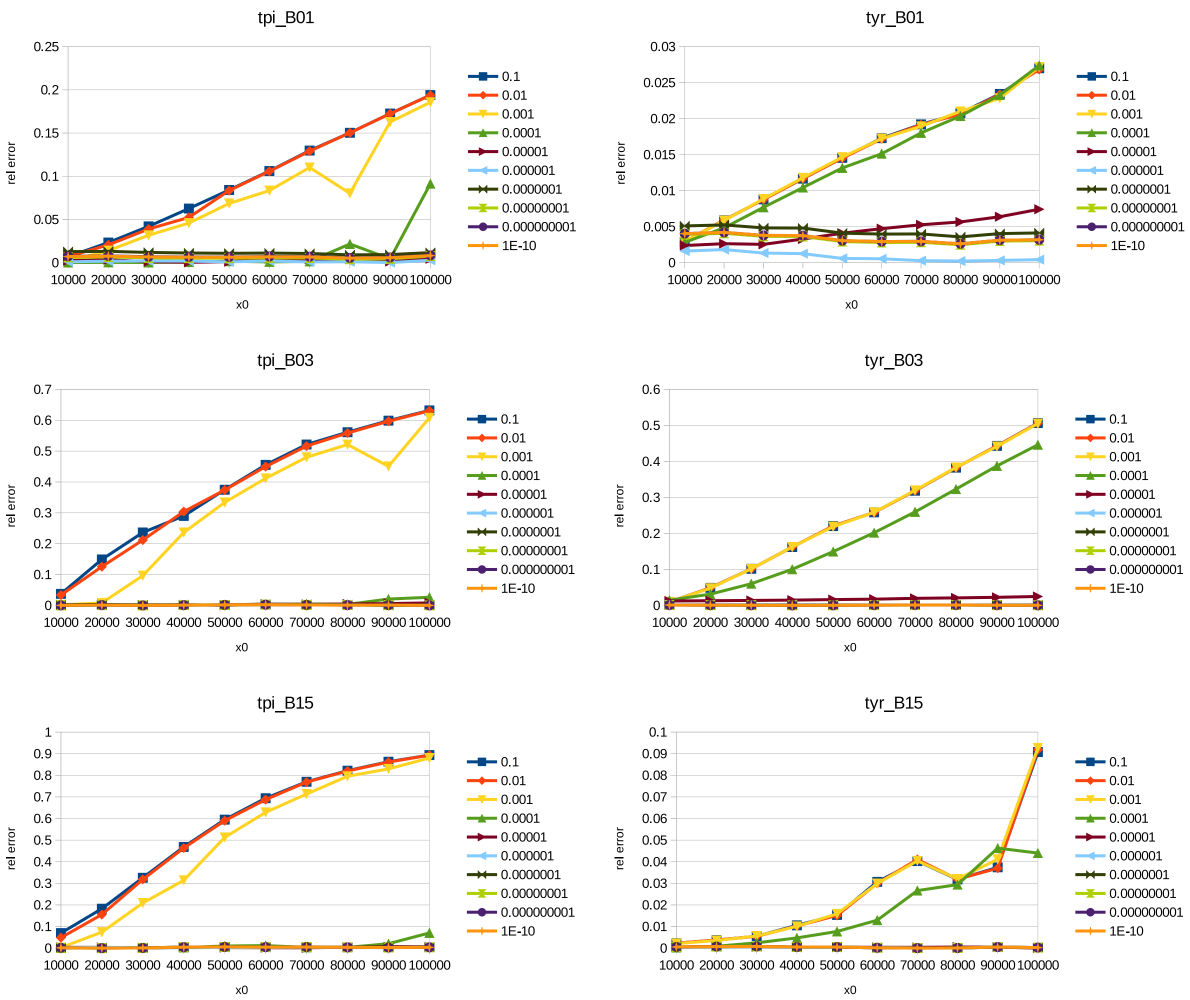

Figure 4.1: Relative errors for a single sample obtained for different values of k (the different lines) and x_0 (horizontal axes), accounting product B01 (top), B03 (middle) and B15 (bottom), using the π estimator $\hat{t}_\pi$ (left) and regression estimator $\hat{t}_{yr}$ (right)

Based on this figure, it is rather clear that the highest values of k tested (10^{-1} to 10^{-4}, corresponding to the curves that diverge from the horizontal axes as the value of x_0 increases) produce inaccurate estimates, compared to the remaining values of k tested. Hence, lower values of k will be more appropriate for this application.

It is, however, less clear from the figure how k values of 10^{-5} and lower compare to each other, as these curves appear to coincide in each of the subplots with comparatively small relative errors. The next step is thus to examine further how the trade off between sample size and accuracy works for each accounting product and various values of k and x_0. In order to fit the tables and since the lowest values of k investigated turn out to produce very similar results, only k values between 10^{-5} through 10^{-8} are included in the tables.

In this step, ten samples are created for each combination of estimator and accounting product and each combination of k and x_0. Both the relative error as well as the sample size as a percentage of the population size are calculated averages over these samples. The standard deviations for the samples as a percentage of the true total technical provisions are also shown in relation to the sample size, to get a sense of the variability between samples. The results of these evaluations can be found in tables B.1 through B.12 in appendix B.

Based on these tables, it is completely clear that $k = 10^{-5}$ and a higher value of x_0 provides estimates that are equally or nearly equally accurate as those obtained with other combinations of k and x_0, while the sample size can be limited to a small fraction of the population in each accounting product. The variability and average relative error do increase somewhat for larger x_0 values, but it appears that $x_0 = 400000$ is still a reasonably suitable choice for all three accounting products. Considering that calculation time varies approximately linearly with the number of policies used in provision calculations, this means calculation times could be cut dramatically.

It is also clear that policies in accounting product B01 (individual occupational pension) have larger surrender values to a greater degree than policies in accounting products B03 or B15 do, since the sample size for B01 as a percentage of the population size is larger for each x_0 tested with $k = 10^{-5}$ compared to the other accounting products. In the portfolio that this thesis considers, B01 has a much smaller number of policies than either B03 or B15, meaning that the total sample size as a percentage of the total number of policies will be much closer to the numbers shown for B03 and B15.

The next step is to run a larger number of simulations using the chosen values $k = 10^{-5}$ and $x_0 = 400000$ to better evaluate the accuracy of the estimators empirically, for each accounting product and estimator type separately. This is done by producing a number of individual samples, for each of which both estimators are calculated. In this investigation, 1 000 samples are created and the same number of estimator calculations have been performed for each accounting product. The results are presented in the following chapter.

Chapter 5

Results and conclusions

The accuracy and variability of the sampling test is indicated in the following tables, where the average estimates for each accounting product and estimation technique are presented along with the true total technical provision, the standard deviation in currency units and as a percentage of the average estimate and an approximate 95% confidence interval for the true aggregate technical provisions.

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 676 018 451	2 676 378 096	48 114 215	1.80%
B03	8 128 171 267	8 123 513 118	117 992 832	1.45%
B15	6 220 145 565	6 222 314 605	97 939 082	1.57%

Table 5.1: Results of the repeated sample simulation, π estimator

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 676 018 451	2 676 213 758	10 021 995	0.37%
B03	8 128 171 267	8 129 267 078	31 885 529	0.39%
B15	6 220 145 565	6 219 784 986	15 679 144	0.25%

Table 5.2: Results of the repeated sample simulation, regression estimator

Accounting	True technical	π estimator		Regression	estimator
product	provision	CI_L	CI_U	CI_L	CI_U
B01	2 676 018 451	2 582 074 235	2 770 681 958	2 656 570 647	2 695 856 868
B03	8 128 171 267	7 892 247 167	8 354 779 070	8 066 771 442	8 191 762 715
B15	6 220 145 565	6 030 354 003	6 414 275 207	6 189 053 863	6 250 516 110

Table 5.3: Approximate 95% confidence intervals for the true aggregate technical provisions

To verify the normality of the estimators, QQ plots have been constructed as shown below:

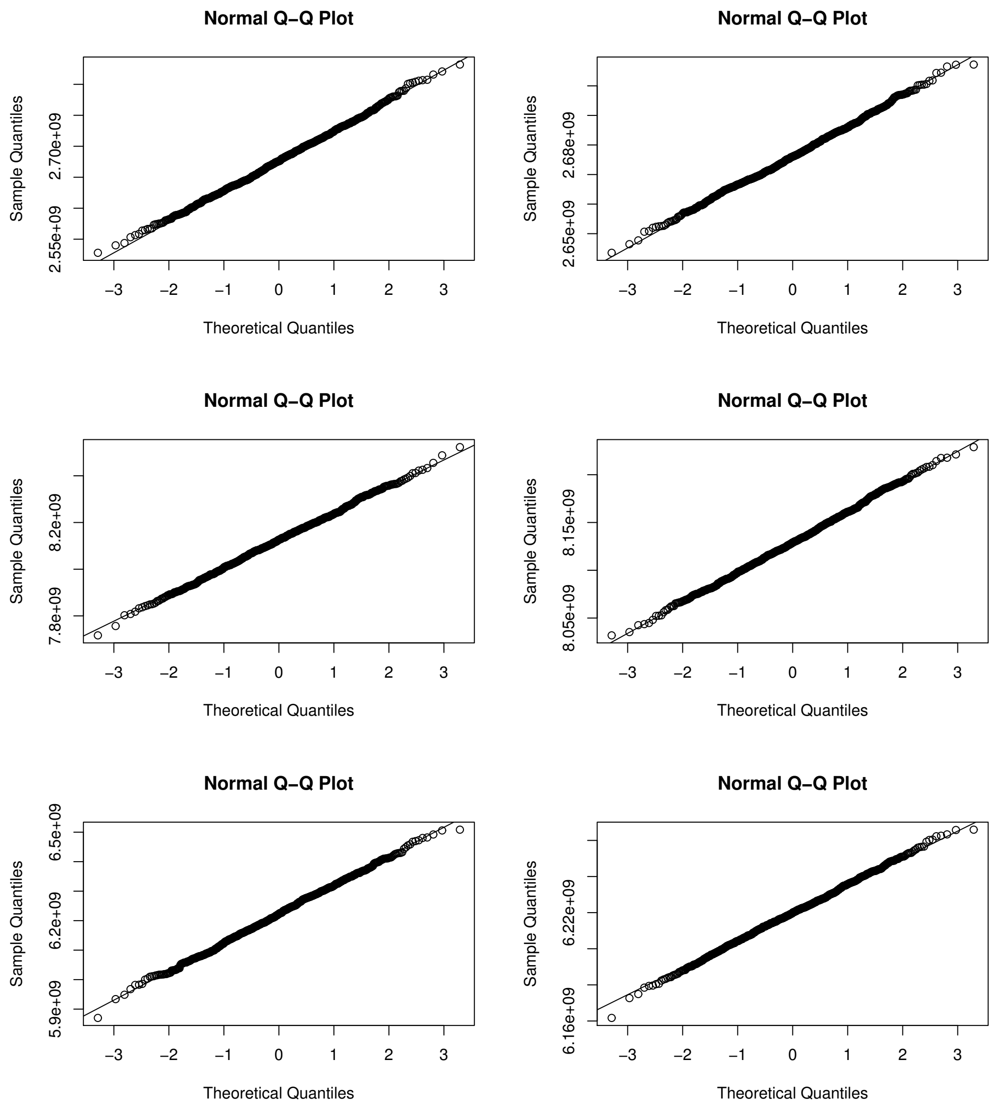

Figure 5.1: QQ plots for the normal distribution, accounting product B01 (top), B03 (middle) and B15 (bottom), using the π estimator $\hat{t}_{\pi}$ (left) and regression estimator $\hat{t}_{yr}$ (right)

The fit to the normal distribution appears quite good in each of these cases, with all points falling very close to the line in each of the subplots.

5.1 Robustness test

As an added test of the robustness of the method for determining inclusion probabilities, the specific k and x_0 parameters chosen for the logistic function and the regression model, the same estimation procedure is performed on data corresponding to three different additional quarters. While the independent variables of the regression model remain the same, separate coefficient estimates are computed for each sample and each quarter. The results are shown in the tables below:

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 460 767 328	2 461 081 510	46 559 658	1.89%
B03	8 191 486 141	8 193 965 337	116 725 302	1.42%
B15	5 993 942 338	5 994 418 731	91 780 339	1.53%

Table 5.4: Results of the repeated sample simulation for Q4 2015, π estimator

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 460 767 328	2 461 040 797	9 946 385	0.40%
B03	8 191 486 141	8 190 359 927	29 232 823	0.36%
B15	5 993 942 338	5 994 095 150	14 681 136	0.24%

Table 5.5: Results of the repeated sample simulation for Q4 2015, regression estimator

Accounting	True technical	π estimator		Regression	estimator
product	provision	CI_L	CI_U	CI_L	CI_U
B01	2 460 767 328	2 369 824 580	2 552 338 441	2 441 545 882	2 480 535 711
B03	8 191 486 141	7 965 183 745	8 422 746 930	8 133 063 593	8 247 656 261
B15	5 993 942 338	5 814 529 268	6 174 308 195	5 965 320 123	6 022 870 177

Table 5.6: Approximate 95% confidence intervals for the true aggregate technical provisions for Q4 2015

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 664 074 513	2 662 519 192	50 588 363	1.90%
B03	8 495 557 921	8 501 920 746	113 967 107	1.34%
B15	6 414 022 584	6 413 520 384	98 998 979	1.54%

Table 5.7: Results of the repeated sample simulation for Q2 2016, π estimator

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 664 074 513	2 664 026 712	10 262 179	0.39%
B03	8 495 557 921	8 497 395 220	37 470 534	0.44%
B15	6 414 022 584	6 414 232 425	16 922 232	0.26%

Table 5.8: Results of the repeated sample simulation for Q2 2016, regression estimator

Accounting	True technical	π estimator		Regression	estimator
product	provision	CI_L	CI_U	CI_L	CI_U
B01	2 664 074 513	2 563 366 001	2 761 672 382	2 643 912 842	2 684 140 583
B03	8 495 557 921	8 278 545 217	8 725 296 275	8 423 952 974	8 570 837 466
B15	6 414 022 584	6 219 482 386	6 607 558 383	6 381 064 851	6 447 399 999

Table 5.9: Approximate 95% confidence intervals for the true aggregate technical provisions for Q2 2016

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 695 738 011	2 694 118 433	50 770 337	1.88%
B03	8 393 618 275	8 399 771 043	112 918 332	1.34%
B15	6 434 549 896	6 433 949 395	99 407 205	1.55%

Table 5.10: Results of the repeated sample simulation for Q3 2016, π estimator

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 695 738 011	2 695 706 454	9 719 144	0.36%
B03	8 393 618 275	8 395 005 954	32 818 961	0.39%
B15	6 434 549 896	6 434 819 514	16 716 947	0.26%

Table 5.11: Results of the repeated sample simulation for Q3 2016, regression estimator

Accounting product	True technical provision	π estimator CI_L	CI_U	Regression estimator CI_L	CI_U
B01	2 695 738 011	2 594 608 573	2 793 628 293	2 676 656 932	2 714 755 976
B03	8 393 618 275	8 178 451 112	8 621 090 974	8 330 680 790	8 459 331 118
B15	6 434 549 896	6 239 111 273	6 628 787 517	6 402 054 297	6 467 584 730

Table 5.12: Approximate 95% confidence intervals for the true aggregate technical provisions for Q3 2016

These earlier quarters appear to be similarly well estimated as the quarter whose data was used to fine tune the estimation method. Hence, the values chosen for k and x_0 and the variables used in the regression model can be expected to perform well over time, without having to recalibrate more than the regression coefficients for each quarterly report.

5.2 Comparison to constant probability sampling

To see whether the approach of unequal probabilities based on policies' surrender values produces any added benefit, a comparison study is performed using a constant selection probability of $\pi = 10\%$, chosen to approximately coincide with the proportion chosen for accounting product B01 with the k and x_0 values discussed above. This produces the following results:

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 676 018 451	2 668 439 768	126 353 690	4.74%
B03	8 128 171 267	8 128 024 172	119 951 100	1.48%
B15	6 220 145 565	6 221 237 879	63 697 141	1.02%

Table 5.13: Results of the repeated sample simulation, π estimator, constant selection probability of 10%

Accounting product	True technical provision	Average (μ)	Standard deviation (σ)	σ/μ (%)
B01	2 676 018 451	2 675 216 703	14 814 616	0.55%
B03	8 128 171 267	8 129 528 293	20 245 336	0.25%
B15	6 220 145 565	6 220 053 318	9 183 191	0.15%

Table 5.14: Results of the repeated sample simulation, regression estimator, constant selection probability of 10%

Accounting	True technical	π estimator		Regression	estimator
product	provision	CI_L	CI_U	CI_L	CI_U
B01	2 676 018 451	2 420 786 535	2 916 093 000	2 646 180 056	2 704 253 350
B03	8 128 171 267	7 892 920 016	8 363 128 328	8 089 847 435	8 169 209 151
B15	6 220 145 565	6 096 391 482	6 346 084 275	6 202 054 263	6 238 052 372

Table 5.15: Approximate 95% confidence intervals for the true aggregate technical provisions, constant selection probability of 10%

Considering that the unequal selection probability approach with the chosen values for k and x_0 produced sample sizes of approximately 3% with comparable variability that is observed in the results above, and that the variability for the x_0 values that result in a sample size closer to 10% (see tables B.1 through B.12 in appendix B) is lower, it does appear that using unequal selection probabilites can offer a better trade off between accuracy/variability and sample size/calculation time than the simpler approach of equal selection probabilities can.

5.3 Conclusions

This thesis has evaluated the use of model assisted survey sampling for the purposes of calculating an accurate estimate of technical provisions of a traditional life insurance portfolio using only a subset of representative policies. The aggregate technical provisions of each accounting product have been estimated using two separate estimation techniques, with the selection probabilities based on the logistic function applied to each policy's surrender value as determined by the source system holding the policy data.

The π estimator uses simple scaling with the help of the selection probabilities, while the regression estimator fits a multiple linear regression model to the sample chosen and uses the same model to explicitly estimate the individual technical provisions of policies not included in the sample.

As seen in tables 5.1 to 5.3, both of these estimation techniques are capable of providing accurate estimates for the aggregate technical provisions of the accounting products considered, and overall the regression estimator exhibits less variance across multiple samples (as should be intuitively expected, considering the good fit obtained within the sample for the regression model chosen).

The selection method used and the range of parameter values tested can produce sample sizes of as little as 3% of the population size (see tables B.1 to B.12) while maintaining a high degree of accuracy with relatively limited variability. As shown in table 5.3, calculated confidence intervals are reasonably narrow and cover the true technical provisions for each accounting product. Since the calculation time in the provision model is approximately linear in the number of policies included, this estimation method allows for highly significant time savings.

The results obtained for the main test data of the last quarter of 2016 can also be expected to hold for other quarters without the need to recalibrate the parameters for the logistic function or to modify the variables used for the regression estimator. Tables 5.4 through 5.12 indicate that accuracy and variability remain comparable over time.

Using unequal selection probabilities does appear to provide an added benefit to the trade off between calculation time and accuracy. Tables 5.13 through 5.15 compared to tables B.1 through B.12 show that for comparable accuracy and variability, the approach with unequal selection probabilities produces smaller sample size, while a comparable sample size produces lower variability for the approach with unequal selection probabilities.

In summary, the sampling methods investigated in this thesis are capable of solving the problem of long calculation times, while still obtaining estimates of the technical provisions that are quite close to the true values determined by applying the provision model to the full portfolio of policies.

5.4 Suggestions for further research

To improve the internal consistency between different regulatory reports, it would be very useful to examine ways to estimate the four main cash flow components (benefit claims, premiums, expenses and taxes) separately for each policy and each monthly period, to ensure that cash flows aggregated for specific purposes do in fact produce the estimated technical provisions when the relevant interest rate curve is applied to calculate the net present value of the cash flows.

The claims component should be possible to predict using the duration and frequency of benefit payments and the calculated benefit per payment as supplied in the data provided to the model. Premiums will be estimated as zero for all policies without an explicit condition of recurring premium payments and should be possible to estimate using the variable for premium payments in the input data for policies with such a condition. As discussed in section A.1, expenses and taxes are calculated from the claims and premium cash flows and technical provisions at each point in time using deterministic administrative loadings, meaning that these would be easily calculated if estimates for claims and premiums have already been determined.

Appendix A

Detailed description of the model

In this appendix, the model is described in more detail, to clarify why the provision calculation process is computationally complex.

A.1 Cash flow model

The technical provisions are contained in a variable called *res_rea_gross_proj*, which contains four cash flow components. Only the expected value of each is modeled, without Monte Carlo simulations. Since this is a traditional life insurance portfolio, there is an embedded guarantee, and only the guaranteed benefit, which is calculated separately and provided as a known amount to the provision model, impacts the technical provisions. For this reason, the shorthand notation for the respective expected values introduced in section 2.1 will be used below.

Cash flow type	Variable name in model	Shorthand notation
Guaranteed claims	*claims_1st_tot*	$c(t)$
Incoming premiums	*premiums_rea_proj*	$p(t)$
Operational expenses	*expense_rea_proj*	$e(t)$
Taxes	*tax_rea_proj*	$x(t)$

Table A.1: Cash flow components of the technical provisions

The guaranteed claims in turn are calculated as the sum of three components:

Claim category	Variable name in model	Shorthand notation
Claims at retirement	*claims_1st_ret*	cr(t)
Claims at death	*claims_1st_dth*	cd(t)
Claims at transfer	*claims_1st_transfer*	ct(t)

Table A.2: Components of the total claims cash flow

Digging into the calculation of these variables, the technical provisions at time t depend on the four cash flow components at time $t+1$ as well as the technical provisions themselves at time $t+1$. Using the shorthand notation from the table, the formula is

$$tp(t) = (c(t+1) - p(t+1) + e(t+1) + x(t+1) + tp(t+1)) \times \frac{df(t+1)}{df(t)},$$

where tp refers to the technical provisions and df to the discount factor from the initial time period. This means there is a recursive dependency only resolved at the end of the projection period, where the provisions and other cash flow components are all zero. The maximum duration considered is

120 years (1 440 months), as policy holders are assumed to not reach ages higher than that.

Furthermore, the operational expense cash flows at time $t+1$ depend on the total claims at time $t+1$, the premiums at time $t+1$ as well as the technical provisions at time $t+1$. The same is true of the tax cash flows at time $t+1$. The figure below illustrates cross-dependencies of the variables in the cash flow model. Here, incoming arrows indicate which values are needed to calculate a certain cash flow component.

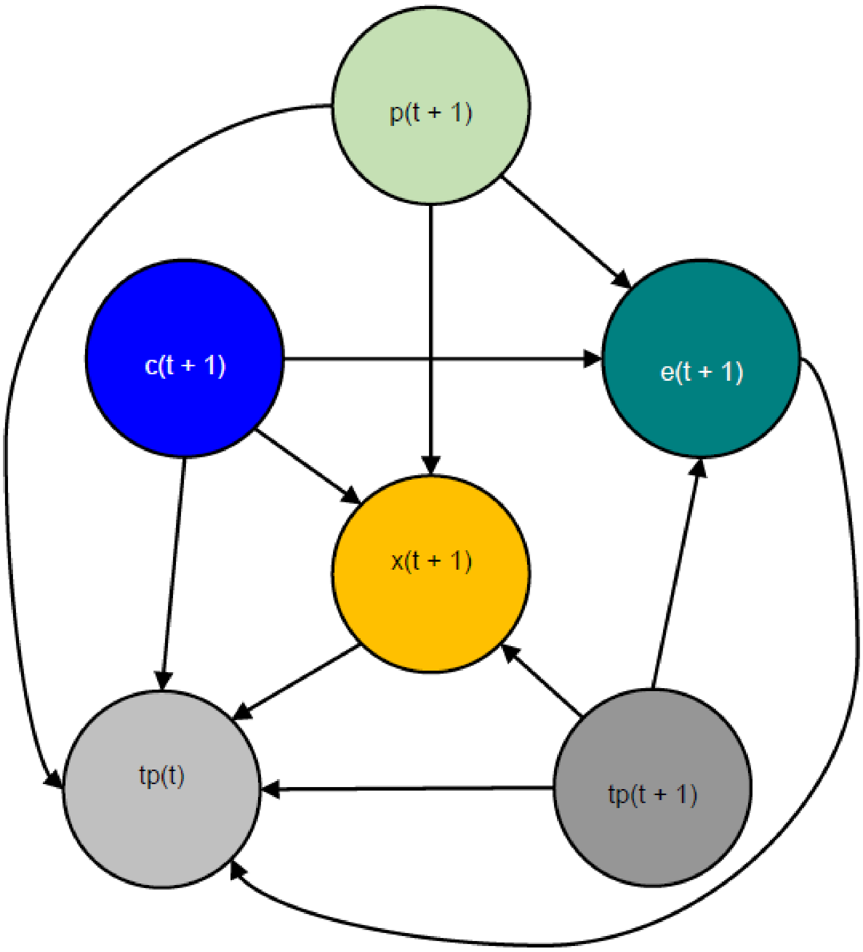

Figure A.1: Cross-dependencies in the cash flow model

Specifically, the formulas for the operational expense and tax cash flows look as follows:

$$\begin{aligned} e(t) = {} & (v-1)\,c(t)\left(1+\frac{l_e}{12}\right)+c(t)\frac{l_e}{12} \\ & -(v\cdot i-1)\,p(t)\left(1+\frac{l_e}{12}\right)-p(t)\frac{l_e}{12} \\ & +tp(t)\left(\frac{l_e}{12}+\frac{1}{2}\left(\frac{l_e}{12}\right)^2\right) \end{aligned}$$

$$x(t)=[v\cdot c(t)-v\cdot i\cdot p(t)]\,\frac{l_t}{12}+tp(t)\left(\frac{l_t}{12}+\frac{l_e\cdot l_t}{144}+\frac{1}{2}\left(\frac{l_t}{12}\right)^2\right).$$

Note here that the operational expense and tax loading parameters (l_e and l_t respectively) are expressed in an annual, continuous manner and should hence be applied to continuous premium and claim cash flows as well as a continuous provision variable. Since the actual provision calculation occurs discretely, the exponential function is approximated by its second order Taylor series. To derive this, note that the sum of operational expenses and taxes in continuous terms over a period of one month is

$$[v \cdot c(t) - v \cdot i \cdot p(t) + tp(t)]\, e^{\frac{l_e + l_t}{12}} - c(t) + p(t) - tp(t),$$

which is approximated by

$$\begin{aligned}
&[v \cdot c(t) - v \cdot i \cdot p(t) + tp(t)] \left(1 + \frac{l_e + l_t}{12} + \frac{1}{2}\left(\frac{l_e + l_t}{12}\right)^2\right) - c(t) + p(t) - tp(t) = \\
&\quad v \cdot c(t) - v \cdot i \cdot p(t) - c(t) + p(t) + [v \cdot c(t) - v \cdot i \cdot p(t) + tp(t)] \frac{l_e + l_t}{12} + \\
&\quad [v \cdot c(t) - v \cdot i \cdot p(t) + tp(t)] \left(\frac{1}{2}\left(\frac{l_e}{12}\right)^2 + \frac{l_e \cdot l_t}{144} + \frac{1}{2}\left(\frac{l_t}{12}\right)^2\right).
\end{aligned}$$

The claim and premium cash flows $c(t)$ and $p(t)$ multiplied by the squared loading parameters in the second order Taylor terms on the last row above will generally be negligible compared to the other terms in the expansion. These terms are thus omitted, in order to reach a somewhat simpler final expression. Reordering the remaining terms so that all terms including l_t are at the end, this becomes

$$\begin{aligned}
&v \cdot c(t) - v \cdot i \cdot p(t) - c(t) + p(t) + [v \cdot c(t) - v \cdot i \cdot p(t) + tp(t)] \frac{l_e}{12} + tp(t)\frac{1}{2}\left(\frac{l_e}{12}\right)^2 + \\
&\quad [v \cdot c(t) - v \cdot i \cdot p(t)] \frac{l_t}{12} + tp(t) \left(\frac{l_t}{12} + \frac{l_e \cdot l_t}{144} + \frac{1}{2}\left(\frac{l_t}{12}\right)^2\right),
\end{aligned}$$

where the terms on the last row correspond to the expression for $x(t)$ given above and the terms on the first row can be further rearranged as

$$\begin{aligned}
&v \cdot c(t) - c(t) + v \cdot c(t)\frac{l_e}{12} - v \cdot i \cdot p(t) + p(t) - v \cdot i \cdot p(t)\frac{l_e}{12} + tp(t)\left(\frac{l_e}{12} + \frac{1}{2}\left(\frac{l_e}{12}\right)^2\right) = \\
&(v - 1)\, c(t)\left(1 + \frac{l_e}{12}\right) + c(t)\frac{l_e}{12} - (v \cdot i - 1)\, p(t)\left(1 + \frac{l_e}{12}\right) - p(t)\frac{l_e}{12} + tp(t)\left(\frac{l_e}{12} + \frac{1}{2}\left(\frac{l_e}{12}\right)^2\right),
\end{aligned}$$

which corresponds to the expression for $e(t)$ presented above. The v and i parameters are administrative loadings on claims and premiums respectively. These are used to allow the company to guarantee a certain fraction of the policy holder's capital while using a certain fraction of capital to seek higher returns through riskier investments, providing discretionary bonus benefits for policy holders.

The premium cash flows are modeled in a simple manner:

$$p(t) = \begin{cases} 0 & \text{for policies without an explicit condition of recurring premiums} \\ 0 & \text{when the end of the premium paying period is reached} \\ l1\_pp(t)\frac{PRMANN}{12} & \text{otherwise,} \end{cases}$$

where $PRMANN$, the expected annual premium, is provided as data to the model (see section A.5) and $l1\_pp(t)$ is the state probability that the policy holder is in state "premium paying" (see section A.2).

The claim cash flows are calculated according to

$$cr(t) = (l1\_popp(t) \times ben\_1st\_pp(t) + l1\_popu(t) \times ben\_1st\_pu(t))\, \frac{m_r}{12 m_m},$$

where $l1\_popp(t)$ and $l1\_popu(t)$ are the state probabilities that the policy holder is receiving payouts (see section A.2), $ben\_1st\_pp(t)$ and $ben\_1st\_pu(t)$ are auxiliary functions that calculate the value of benefits to the policy holder recursively for premium paying and paid-up policies respectively and m_r and m_m are discretization adjustment factors;

$$cd(t) = l2\_po(t) \times ben\_1st\_l2(t)\frac{m_r}{12m_m},$$

where $l2\_po(t)$ is the state probability that the beneficiary is receiving payouts (see section A.2), $ben\_1st\_l2(t)$ is an auxiliary function that calculates the value of benefits to a beneficiary recursively and m_r and m_m are discretization adjustment factors; and

$$ct(t) = (t1\_pp\_tr(t) \times res\_sv\_pp(t) + t1\_pu\_tr(t) \times res\_sv\_pu(t))\,\frac{m_r}{12m_m},$$

where $t1\_pp\_tr(t)$ and $t1\_pu\_tr(t)$ are the state probability increments that the policy holder is transferring the policy (see section A.2), $res\_sv\_pp(t)$ and $res\_sv\_pu(t)$ are auxiliary functions that calculate the surrender value of the policy for premium paying or paid-up respectively (by using relevant commutation factors) and m_r and m_m are discretization adjustment factors.

A.2 State model

The state model uses functions keeping track of the probabilities of being in each of the possible states at each point in time as well as probability increments to transition between states. The former category uses variable names starting with $l1$ for the policy holder and $l2$ for the possible beneficiary in the case the policy holder dies, respectively. The latter category similarly uses variable names starting with $t1$ and $t2$. Below are the formulas for these variables, with some comments to illustrate certain features.

Note specifically that the functions are *not* structured in the typical matrix algebra sense described in section 2.1; there is no explicit matrix formed for transition probabilities and in fact only increments to state probabilities are calculated. The relationship between the ordinary multiplicative approach and this alternative additive approach can be derived as follows:

Consider a particular state x. The probability of being in this state at time $t+1$ is

$$\begin{aligned}
p_x(t+1) = &\sum_{i=1}^{N} p_i(t)p_{ix}(t) \\
&= p_1(t)p_{1x}(t) + p_2(t)p_{2x}(t) + \ldots + p_x(t)p_{xx}(t) + \ldots + p_N(t)p_{Nx}(t) \\
&= p_x(t) + p_1(t)p_{1x}(t) + p_2(t)p_{2x}(t) + \ldots + p_{x-1}(t)p_{x-1,x}(t) + \\
&\quad p_{x+1}(t)p_{x+1,x}(t) + \ldots + p_N(t)p_{Nx}(t) - p_x(t)\,(1 - p_{xx}(t))\,.
\end{aligned}$$

By introducing the notation $t_{ix}(t) = p_i(t)p_{ix}(t)$, this can be rewritten as

$$p_x(t+1) = p_x(t) + t_{1x}(t) + t_{2x}(t) + \ldots + t_{x-1,x}(t) + t_{x+1,x}(t) + \ldots + t_{Nx}(t) - p_x(t)\,(1 - p_{xx}(t))\,,$$

where the $t_{ix}(t)$ can be thought of as probability increments for the transitions out of each state i into state x.

For an absorbing state x, $1 - p_{xx}(t) = 0$, and hence the probability of being in such a state at time $t+1$ is simply the probability of being in the same state at time t plus the sum of the probability increments for transitioning into state x.

For a non-absorbing state x, it is more helpful to think of the term $p_x(t)\,(1 - p_{xx}(t))$ as the sum of probability increments for transitions out of state x. The correctness of this alternative interpretation can be seen from the fact that the sum of probabilities of transitioning from each state x must equal one:

$$\sum_{j=1}^{N} p_{xj}(t) = 1 \Leftrightarrow 1 - p_{xx}(t) = p_{x1}(t) + p_{x2}(t) + \ldots + p_{x,x-1}(t) + p_{x,x+1}(t) + \ldots + p_{xN}(t),$$

and hence the probability of being in state x at time $t+1$ can be rewritten as

$$\begin{aligned} p_x(t+1) = p_x(t) \; & + t_{1x}(t) + t_{2x}(t) + \ldots + t_{x-1,x}(t) + t_{x+1,x}(t) + \ldots + t_{Nx}(t) \\ & - t_{x1}(t) - t_{x2}(t) - \ldots - t_{x,x-1}(t) - t_{x,x+1}(t) - \ldots - t_{x,N}(t) \end{aligned}$$

in this case. In the sections below, the expressions for state probabilities and transition probability increments used in the provision model are provided.

A.2.1 State probabilities

Probability that the policy holder is in absorbing state "DE" (dead) at time t (probability that the policy holder was in state "DE" at time $t-1$ plus the probability increments that the policy holder transitioned to state "DE" from the states that can transition there):

$$l1\_de(t) = l1\_de(t-1) + t1\_pp\_de(t) + t1\_pu\_de(t) + t1\_popp\_de(t) + t1\_popu\_de(t)$$

Probability that the policy holder is in absorbing state "MA" (matured) at time t (probability that the policy holder was in state "MA" at time $t-1$ plus the probability increments that the policy holder transitioned to state "MA" from the states that can transition there):

$$l1\_ma(t) = l1\_ma(t-1) + t1\_pp\_ma(t) + t1\_pu\_ma(t) + t1\_popp\_ma(t) + t1\_popu\_ma(t)$$

Probability that the policy holder is in state "POPP" (paying out, from premium paying) at time t (probability that the policy holder was in state "POPP" at time $t-1$ plus the probability increment that the policy holder transitioned to state "PO" from "PP" minus the probability increments that the policy holder transitioned from state "POPP" to "MA", "DE" or "TR"):

$$l1\_popp(t) = l1\_popp(t-1) + t1\_pp\_po(t) - t1\_popp\_ma(t) - t1\_popp\_de(t) - t1\_popp\_tr(t)$$

Probability that the policy holder is in state "POPU" (paying out, from paid-up) at time t (probability that the policy holder was in state "POPU" at time $t-1$ plus the probability increment that the policy holder transitioned to state "PO" from "PU" minus the probability increments that the policy holder transitioned from state "POPU" to "MA", "DE" or "TR"):

$$l1\_popu(t) = l1\_popu(t-1) + t1\_pu\_po(t) - t1\_popu\_ma(t) - t1\_popu\_de(t) - t1\_popu\_tr(t)$$

Probability that the policy holder is in state "PP" (premium paying) at time t (probability that the policy holder was in state "PP" at time $t-1$ minus the probability increments that the policy holder transitioned from state "PP" to the other possible states):

$$l1\_pp(t) = l1\_pp(t-1) - t1\_pp\_de(t) - t1\_pp\_ma(t) - t1\_pp\_po(t) - t1\_pp\_pu(t) - t1\_pp\_tr(t)$$

Probability that the policy holder is in state "PU" (paid-up) at time t (probability that the policy holder was in state "PU" at time $t-1$ plus the probability increment that the policy holder transitioned from "PP" to "PU" minus the probability increments that the policy holder transitioned from state "PU" to the other possible states):

$$l1\_pu(t) = l1\_pu(t-1) + t1\_pp\_pu(t) - t1\_pu\_de(t) - t1\_pu\_ma(t) - t1\_pu\_po(t) - t1\_pu\_tr(t)$$

Probability that the policy holder is in absorbing state "TR" (transferred) at time t (probability that the policy holder was in state "TR" at time $t-1$ plus the probability increments that the policy holder transitioned to state "TR" from the states that can transition there):

$$l1\_tr(t) = l1\_tr(t-1) + t1\_pp\_tr(t) + t1\_pu\_tr(t) + t1\_popp\_tr(t) + t1\_popu\_tr(t)$$

In all cases, these formulas show the values of the state probabilities at an arbitrary point in time after the initial state. The policy data used in the calculation indicates what state a given policy is in at the start of the calculation. The model assigns 1 as the probability of that state at time 0 and zero probability for each of the other states.

Specifically for the policies with a beneficiary if the policy holder dies, the cash flow model requires a further state $l2\_po$ to keep track of whether the policy is paying to the beneficiary (the terms added indicate whether the policy holder has transitioned to state "DE"; the calculation does *not* take into account the possibility that the beneficiary also might be dead):

$$l2\_po(t) = l2\_po(t-1) + t1\_pp\_de(t) + t1\_pu\_de(t) + t1\_popp\_de(t) + t1\_popu\_de(t)$$

A.2.2 Transition probability increments

Probability increment that the policy holder transitions from state "POPP" to state "DE" at time t (probability that the policy holder was in state "POPP" at time $t-1$, minus the probability increment that the policy holder transitioned to state "MA" at time t because maturity takes precedence over death, multiplied by the mortality rate):

$$t1\_popp\_de(t) = [l1\_popp(t-1) - t1\_popp\_ma(t)] \times mort\_rate\_exp(t)$$

Probability increment that the policy holder transitions from state "POPP" to state "MA" at time t (the raw policy data contains the date of maturity, allowing the period of maturity to be easily determined):

$$t1\_popp\_ma(t) = \begin{cases} l1\_popp(t-1) & \text{if } t \text{ is the period of maturity} \\ 0 & \text{otherwise} \end{cases}$$

Probability increment that the policy holder transitions from state "POPP" to state "TR" at time t (probability that the policy holder was in state "POPP" at time $t-1$, minus the probability increments that the policy holder transitioned to state "MA" or "DE" at time t because maturity and death take precedence over transfer, multiplied by the surrender rate):

$$t1\_popp\_tr(t) = [l1\_popp(t-1) - t1\_popp\_ma(t) - t1\_popp\_de(t)] \times surrender\_rate(t)$$

Probability increment that the policy holder transitions from state "POPU" to state "DE" at time t (probability that the policy holder was in state "POPU" at time $t-1$, minus the probability increment that the policy holder transitioned to state "MA" at time t because maturity takes precedence over death, multiplied by the mortality rate):

$$t1\_popu\_de(t) = [l1\_popu(t-1) - t1\_popu\_ma(t)] \times mort\_rate\_exp(t)$$

Probability increment that the policy holder transitions from state "POPU" to state "MA" at time t (the raw policy data contains the date of maturity, allowing the period of maturity to be easily determined):

$$t1\_popu\_ma(t) = \begin{cases} l1\_popu(t-1) & \text{if } t \text{ is the period of maturity} \\ 0 & \text{otherwise} \end{cases}$$

Probability increment that the policy holder transitions from state "POPU" to state "TR" at time t (probability that the policy holder was in state "POPU" at time $t-1$, minus the probability increments that the policy holder transitioned to state "MA" or "DE" at time t because maturity and death take precedence over transfer, multiplied by the surrender rate):

$$t1\_popu\_tr(t) = [l1\_popu(t-1) - t1\_popu\_ma(t) - t1\_popu\_de(t)] \times surrender\_rate(t)$$

Probability increment that the policy holder transitions from state "PP" to state "DE" at time t (probability that the policy holder was in state "PP" at time $t-1$, minus the probability increment that the policy holder transitioned to state "MA" at time t because maturity takes precedence over death, multiplied by the mortality rate):

$$t1\_pp\_de(t) = [l1\_pp(t-1) - t1\_pp\_ma(t)] \times mort\_rate\_exp(t)$$

Probability increment that the policy holder transitions from state "PP" to state "MA" at time t (the raw policy data contains the date of maturity, allowing the period of maturity to be easily determined):

$$t1\_pp\_ma(t) = \begin{cases} l1\_pp(t-1) & \text{if } t \text{ is the period of maturity} \\ 0 & \text{otherwise} \end{cases}$$

Probability increment that the policy holder transitions from state "PP" to state "PO" at time t (probability that the policy holder was in state "PP" at time $t-1$, minus the probability increments that the policy holder transitioned to the other possible states):

$$t1\_pp\_po(t) = l1\_pp(t-1) - t1\_pp\_pu(t) - t1\_pp\_ma(t) - t1\_pp\_de(t) - t1\_pp\_tr(t)$$

Probability increment that the policy holder transitions from state "PP" to state "PU" at time t (probability that the policy holder was in state "PP" at time $t-1$, minus the probability increments that the policy holder transitioned to state "MA" or "DE" at time t because maturity and death take precedence over transfer, multiplied by 1 minus half the surrender rate because premium cessation and transfers are assumed to occur simultaneously, further multiplied by the paid-up rate):

$$t1\_pp\_pu(t) = [l1\_pp(t-1) - t1\_pp\_ma(t) - t1\_pp\_de(t)] \times \left[1 - \frac{surrender\_rate(t)}{2}\right] \times paidup\_rate(t)$$

Probability increment that the policy holder transitions from state "PP" to state "TR" at time t (probability that the policy holder was in state "PP" at time $t-1$, minus the probability increments that the policy holder transitioned to state "MA" or "DE" at time t because maturity and death take precedence over transfer, multiplied by 1 minus half the paid-up rate because premium cessation and transfers are assumed to occur simultaneously, further multiplied by the surrender rate):

$$t1\_pp\_tr(t) = [l1\_pp(t-1) - t1\_pp\_ma(t) - t1\_pp\_de(t)] \times \left[1 - \frac{paidup\_rate(t)}{2}\right] \times surrender\_rate(t)$$

Note that the above two expressions mean that the sum of the transition probability increments from "PP" to "PU" and "TR" becomes

$$\begin{aligned} t1\_pp\_pu(t) + t1\_pp\_tr(t) = {} & [l1\_pp(t-1) - t1\_pp\_ma(t) - t1\_pp\_de(t)] \times \\ & [paidup\_rate(t) + surrender\_rate(t) - \\ & paidup\_rate(t) \times surrender\_rate(t)], \end{aligned}$$

where the product of the paid-up rate and the surrender rate is subtracted to not double count policies, as a given policy that transfers out of the company also will cease premium payments.

Probability increment that the policy holder transitions from state "PU" to state "DE" at time t (probability that the policy holder was in state "PU" at time $t-1$, minus the probability increment that the policy holder transitioned to state "MA" at time t because maturity takes precedence over death, multiplied by the mortality rate):

$$t1\_pu\_de(t) = [l1\_pu(t-1) - t1\_pu\_ma(t)] \times mort\_rate\_exp(t)$$

Probability increment that the policy holder transitions from state "PU" to state "MA" at time t (the raw policy data contains the date of maturity, allowing the period of maturity to be easily determined):

$$t1\_pu\_ma(t) = \begin{cases} l1\_pu(t-1) & \text{if } t \text{ is the period of maturity} \\ 0 & \text{otherwise} \end{cases}$$

Probability increment that the policy holder transitions from state "PU" to state "PO" at time t (probability that the policy holder was in state "PU" at time $t-1$, plus the probability increment that the policy holder transitioned from "PP" to "PU", minus the probability increments that the policy holder transitioned to the other possible states):

$$t1\_pu\_po(t) = l1\_pu(t-1) + t1\_pp\_pu(t) - t1\_pu\_ma(t) - t1\_pu\_de(t) - t1\_pu\_tr(t)$$

Probability increment that the policy holder transitions from state "PU" to state "TR" at time t (probability that the policy holder was in state "PU" at time $t-1$, minus the probability increments that the policy holder transitioned to state "MA" or "DE" at time t because maturity and death take precedence over transfer, multiplied by the surrender rate):

$$t1\_pu\_tr(t) = [l1\_pu(t-1) - t1\_pu\_ma(t) - t1\_pu\_de(t)] \times surrender\_rate(t)$$

A.3 Parameter sets

The model uses a range of different parameters, which are organised as follows:

Product group: Collectively negotiated occupational pension insurance uses a different operational expense loading compared to individual occupational pension insurance and other life insurance, reflecting different actual operational expenses for these lines of business. Occupational pension insurance also uses a separate interest rate curve for discounting purposes compared to other life insurance, as required by the regulator. Furthermore, other life insurance uses "safe" assumptions regarding mortality, while occupational pension insurance uses "realistic" assumptions.

Tax classification: The Swedish tax code separates life insurance into tax classes P ("pension insurance") and K ("capital insurance") and these are taxed at different rates. Furthermore, while both classes apply the tax rate to the yield of government bonds of approximately 10 year durations with a suitable adjustment, different adjustments are made to the yield for the two classes. Hence, the tax loading (used for setting aside funds from policies to pay their taxes) is separate for P and K.

It should be noted that collectively negotiated policies are only classified as occupational pension insurance and only as tax class P, while individual policies can belong to either of the four possible combinations of occupational pension/other life insurance and tax class P/K. Hence, there is a total of five possible such combinations, referred to as *accounting products*. Below is a table of the accounting products included in the portfolio studied in this thesis:

Accounting product	Contents
B01	Individual occupational pension insurance, tax class P
B02	Individual occupational pension insurance, tax class K
B03	Other life insurance, tax class P
B04	Other life insurance, tax class K
B15	Collectively negotiated occupational pension insurance, tax class P

Table A.3: Accounting products present in the portfolio studied

Some parameter values are separate for each accounting product, for instance parameters used for valuing the policy holder option (available in some but not all policies) of changing the length of the benefit payment period. These parameters include the probability of changing benefit duration as well as the expected change in this duration, and are further split between policies that initially had life-long benefit duration and policies that initially had a fixed maximum benefit duration.

Technical basis: Sets of parameter values applicable to policies issued on or after a certain starting date. Typically, there are several simultaneously active technical bases for each of the tax classes and for different collectives of policy holders. Occasionally, new technical bases are introduced in response to changing financial market conditions and their effects on the investment portfolio.

Newly issued policies will then use a relevant new basis.

For policies without an explicit condition of recurring premiums, existing capital keeps its original technical basis while any further premiums paid are placed in sub-policies that will be subject to the technical bases applicable at the time of payment. A policy can also be issued with an explicit condition of recurring premium payments, in which case the original technical basis is applied for all future premiums as long as premium payments are made according to the condition.

Mortality basis: A set of Makeham parameters for calculating the survival function and commutation factors. Each technical basis can contain separate mortality bases for purposes such as provision calculations, pricing, prognoses etc. Several technical bases can share the same mortality basis for one or several of these purpose classes.

To further illustrate the organization of the parameter sets, below are sample tables of fictitious (but realistic in terms of the order of magnitude) parameter values:

	196400K	196400P	196401K	196401P	199300K	199300P	...
mortality basis pricing	mort1	mort2	mort1	mort2	mort3	mort4	...
mortality basis provisions	mort5	mort5	mort5	mort5	mort6	mort6	...
mortality basis prognosis	mort7	mort8	mort9	mort10	mort11	mort12	...
mortality basis realistic	real3	real3	real3	real3	real3	real3	...
survival basis pricing	surv1	surv2	surv1	surv2	surv3	surv4	...
survival basis provisions	surv5	surv5	surv5	surv5	surv6	surv6	...
survival basis prognosis	surv7	surv8	surv9	surv10	surv11	surv12	...
survival basis realistic	real3	real3	real3	real3	real3	real3	...
delta pricing	0.0123	0.0157	0.0123	0.0157	0.0214	0.0214	...
delta provisions	0.0234	0.0178	0.0234	0.0178	0.0235	0.0235	...
delta prognosis	0.0184	0.0125	0.0184	0.0125	0.0209	0.0209	...

Table A.4: Sample, fictitious table of technical bases

As indicated in the table, a particular mortality basis can be shared by several technical bases, but can also be unique to a single technical basis. The numbers at the start of the technical basis names indicate the year it was introduced as well as identifier numbers in case several bases were introduced the same year. The last letter reflects the tax class - the two tax classes can have different technical bases.

	mort1_M	mort1_F	mort2_M	...	surv1_M	surv1_F	...
α	0.0015	0.0015	0.0018	...	0.0025	0.0025	...
β	0.000057	0.000057	0.000061	...	0.000043	0.000043	...
γ	0.051	0.051	0.047	...	0.043	0.043	...
f	0	3	0	...	0	4	...

Table A.5: Sample, fictitious table of mortality bases

The names of the mortality bases reflect the names used in the table for technical bases above and the final letter corresponds to the gender of the policy holder. Some mortality bases are gender neutral for legal reasons, in which case the _M and _F versions will have the same Makeham parameters.

A.4 Calculation of commutation factors

To calculate the survival function $l(x)$, we note the fact that the Makeham model uses the mortality intensity $\mu_x = \alpha + \beta e^{\gamma(x-f)}, x \geq 0$ and insert this into the general expression linking the mortality intensity and the survival function:

$$l(x) = exp\left(-\int_0^x \mu_s ds\right) = exp\left(-\int_0^x \left(\alpha + \beta e^{\gamma(s-f)}\right) ds\right) = exp\left[-\alpha x - \frac{\beta}{\gamma}\left(e^{\gamma(x-f)} - 1\right)\right]$$

In this model, $l(x)$ is considered to be zero for $x > 120$ years, and for $x > 80$ years the mortality intensity is considered to increase linearly rather than exponentially.

The $D(x)$ commutation factor is trivially calculated according to its definition, $D(x) = l(x)e^{-\delta x}$, though the δ value can vary by technical basis and not simply by mortality basis as seen in the previous section. For the purposes of calculating the $N(x)$ commutation factor, Simpson's rule is used to approximate the value of the integral. This means that a general definite integral is approximated by

$$\int_a^b f(x)dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right]$$

The approximate $N(x)$ values for a given age x (expressed in months) are calculated cumulatively, by estimating such integral values over the required number of two-month intervals and adding them up, according to the following recursive formula:

$$N(x) = \frac{D(x) + 4D(x+1) + D(x+2)}{36} + N(x+2)$$

As a consequence of the duration changes according to the option available to some policy holders, $N(x)$ commutation factors for non-integer ages in months can be required in provision calculations. These are then calculated using a Taylor expansion based interpolation between the values available in this table.

The original model calculates commutation factors on an as-needed basis and stores them in an expanding but non-persistent table to be re-used by any further policies sharing the same parameters. For subsequent quarters, the entire table is recalculated. This method is flexible and allows further technical bases and/or mortality bases to be introduced without requiring modifications to the method or separate preparatory calculations of the commutation factors. To save time in the main technical provision calculation loop of the adapted R version of the model, all possible commutation factor values for each combination of technical basis, mortality basis and age (measured in months) have been calculated separately and stored in tables which are loaded at the start of the main program.

A.5 Format of policy data

The majority of information required to actually calculate technical provisions is supplied to the model by the data from the policy information database. The following table shows the variables included in this data and brief descriptions of their content:

Variable name	Contents
POLICY_ID	A unique identifier for each policy
ISSUE_YEAR	The year the policy was issued
ISSUE_MTH	The month the policy was issued
ISSUE_DAY	The day the policy was issued
DOB1_YEAR	The year the policy holder was born
DOB1_MTH	The month the policy holder was born
DOB1_DAY	The day the policy holder was born
SEX1	The gender of the policy holder (M/F)
AGE_T_TYPE1	Technical age type of policy holder ("E" for exact or "H" for whole years)
PROD_GROUP	"Ind" for individually signed policies, "TJP" for collectively negotiated
TJP_CODE	"1" for occupational pension insurance, "0" for other life insurance
TARIFF	The tariff type of the policy (see section 2.4)
SAR_IND1	An index for whether the policy has mortality (0) or longevity (-1) risk
TECH_BASIS	The technical basis of the policy/subpolicy (see section A.3)
TAX_CLASS	The tax class of the policy (see section A.3)
PRM_FLAG	"L" for policies with an explicit condition of recurring premiums, "E" for purely one-off premium, "LE" otherwise
STATUS	The state in the state model at time 0 (see section 2.3)
PRE_AGE_AC	Policy holder age when premium payments are scheduled to cease
RET_AGE_TE	Technical retirement age of policy holder
RET_AGE_AC	Actual retirement age of policy holder (may be different from RET_AGE_TE if AGE_T_TYPE1 = "H")
BEN_TRM_AC	Length of benefit payment period, policy holder
FAM_TRM_AC	length of benefit payment period, beneficiary (if any)
BEN_1ST	Calculated guaranteed benefit per payment
BEN_FREQ	Number of benefit payments per year
BEN_TIME	Benefit timing ("B" for advance, "A" for arrears)
PRMANN	Annual premium at calculation date (zero if PRM_FLAG $\neq$ "L")
PRMTRM	Premium per period at calculation date (zero if PRM_FLAG $\neq$ "L")
PRM_FREQ	Number of premiums per year (zero if PRM_FLAG $\neq$ "L")
SURR_FL	"J" if policy can be transferred, otherwise "N"
SV1ST	Surrender value of policy if transferred out of the company
RES_1ST	Premium reserve of policy
RES_2ND	Second order reserve of policy (including discretionary bonus)

Table A.6: Contents of data used by the model

Note specifically that an estimate of the guaranteed benefit of each policy is calculated beforehand and provided to the technical provision model under the variable name BEN_1ST. The model does contain functions to recalculate benefits in the event of premium cessation however.

Appendix B

Accuracy and variability of evaluated logistic function parameters

Below are tables showing the accuracy and variability of the π estimator and the regression estimator for each of the accounting products considered and for the values of k and x_0 evaluated. The accuracy is measured as the average of the relative error observed for the aggregate technical provision estimate in each of the 10 samples chosen for each case. The variability is measured as the standard deviation as a percentage of the true aggregate technical provisions for the 10 samples used for each estimator and combination of k and x_0.

Note that since both the π estimator and the regression estimator are unbiased per Särndal et al [8], if one instead calculates the average aggregate technical provisions and then calculates the relative error of this average, the result will be a smaller error than the average of the individual sample relative errors reflected in the tables below.

B.1 The π estimator

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.27%	62.7%	0.63%	52.2%	1.10%	50.2%	0.92%	50.0%
40000	0.28%	58.6%	0.64%	51.8%	1.00%	50.2%	0.89%	50.0%
60000	0.21%	54.4%	0.66%	51.3%	1.01%	50.1%	0.94%	50.0%
80000	0.20%	50.2%	0.63%	50.8%	0.92%	50.1%	0.87%	50.0%
100000	0.29%	46.1%	0.63%	50.3%	0.97%	50.0%	0.90%	49.9%
120000	0.20%	42.2%	0.61%	49.8%	0.97%	50.0%	0.92%	49.9%
140000	0.19%	38.5%	0.79%	49.5%	0.94%	50.0%	0.85%	50.0%
160000	0.44%	34.9%	0.73%	49.0%	0.91%	50.0%	0.79%	50.1%
180000	0.32%	31.4%	0.77%	48.5%	0.98%	49.9%	0.82%	50.0%
200000	0.23%	28.5%	0.81%	48.0%	0.95%	49.9%	0.81%	50.0%
220000	0.35%	25.7%	0.96%	47.5%	1.06%	49.9%	0.92%	50.1%
240000	0.60%	23.2%	1.04%	47.0%	1.06%	49.8%	0.94%	50.1%
260000	0.91%	20.9%	0.96%	46.6%	1.06%	49.8%	0.86%	50.1%
280000	0.86%	18.8%	0.93%	46.1%	1.06%	49.7%	0.83%	50.1%
300000	0.93%	17.0%	0.93%	45.6%	1.08%	49.7%	0.88%	50.1%
320000	0.77%	15.4%	1.09%	45.2%	1.34%	49.7%	1.13%	50.2%
340000	1.06%	13.9%	1.14%	44.7%	1.34%	49.7%	1.15%	50.2%
360000	1.29%	12.6%	1.11%	44.2%	1.41%	49.6%	1.15%	50.2%
380000	1.32%	11.4%	1.23%	43.7%	1.40%	49.6%	1.22%	50.2%
400000	1.11%	10.4%	1.32%	43.2%	1.55%	49.5%	1.29%	50.1%

Table B.1: Accuracy and sample size for the π estimator, accounting product B01

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.36%	62.7%	0.89%	52.2%	1.45%	50.2%	1.23%	50.0%
40000	0.32%	58.6%	0.86%	51.8%	1.40%	50.2%	1.19%	50.0%
60000	0.32%	54.4%	0.90%	51.3%	1.45%	50.1%	1.26%	50.0%
80000	0.28%	50.2%	0.83%	50.8%	1.31%	50.1%	1.12%	50.0%
100000	0.35%	46.1%	0.87%	50.3%	1.28%	50.0%	1.13%	49.9%
120000	0.25%	42.2%	0.76%	49.8%	1.34%	50.0%	1.13%	49.9%
140000	0.26%	38.5%	0.94%	49.5%	1.15%	50.0%	1.02%	50.0%
160000	0.54%	34.9%	0.88%	49.0%	1.18%	50.0%	0.99%	50.1%
180000	0.35%	31.4%	0.98%	48.5%	1.25%	49.9%	1.03%	50.0%
200000	0.27%	28.5%	1.02%	48.0%	1.28%	49.9%	1.06%	50.0%
220000	0.44%	25.7%	1.17%	47.5%	1.33%	49.9%	1.10%	50.1%
240000	0.83%	23.2%	1.22%	47.0%	1.30%	49.8%	1.11%	50.1%
260000	1.06%	20.9%	1.11%	46.6%	1.27%	49.8%	1.04%	50.1%
280000	1.09%	18.8%	0.99%	46.1%	1.25%	49.7%	0.98%	50.1%
300000	1.05%	17.0%	0.98%	45.6%	1.23%	49.7%	0.96%	50.1%
320000	1.01%	15.4%	1.09%	45.2%	1.43%	49.7%	1.17%	50.2%
340000	1.38%	13.9%	1.17%	44.7%	1.51%	49.7%	1.28%	50.2%
360000	1.58%	12.6%	1.09%	44.2%	1.52%	49.6%	1.24%	50.2%
380000	1.59%	11.4%	1.06%	43.7%	1.52%	49.6%	1.27%	50.2%
400000	1.56%	10.4%	1.37%	43.2%	1.74%	49.5%	1.48%	50.1%

Table B.2: Variability and sample size for the π estimator, accounting product B01

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.16%	52.1%	0.32%	50.3%	0.37%	50.0%	0.34%	50.0%
40000	0.15%	47.3%	0.32%	49.8%	0.37%	50.0%	0.37%	50.0%
60000	0.16%	42.6%	0.35%	49.3%	0.39%	49.9%	0.39%	50.0%
80000	0.26%	37.9%	0.27%	48.8%	0.31%	49.9%	0.31%	50.0%
100000	0.23%	33.6%	0.29%	48.3%	0.36%	49.8%	0.35%	50.0%
120000	0.36%	29.4%	0.29%	47.8%	0.34%	49.8%	0.34%	50.0%
140000	0.39%	25.6%	0.23%	47.3%	0.32%	49.7%	0.31%	50.0%
160000	0.40%	22.2%	0.22%	46.8%	0.31%	49.7%	0.30%	50.0%
180000	0.28%	19.1%	0.24%	46.3%	0.30%	49.6%	0.34%	50.0%
200000	0.29%	16.4%	0.32%	45.8%	0.37%	49.6%	0.37%	50.0%
220000	0.40%	14.0%	0.32%	45.3%	0.40%	49.5%	0.41%	49.9%
240000	0.39%	11.9%	0.35%	44.8%	0.43%	49.5%	0.43%	49.9%
260000	0.50%	10.1%	0.27%	44.3%	0.33%	49.4%	0.32%	49.9%
280000	0.62%	8.5%	0.22%	43.8%	0.32%	49.4%	0.31%	49.9%
300000	0.77%	7.2%	0.29%	43.3%	0.40%	49.3%	0.40%	49.9%
320000	0.94%	6.0%	0.29%	42.8%	0.40%	49.3%	0.39%	49.9%
340000	1.27%	5.1%	0.35%	42.3%	0.43%	49.2%	0.42%	49.9%
360000	1.04%	4.3%	0.34%	41.8%	0.41%	49.2%	0.40%	49.9%
380000	1.17%	3.6%	0.39%	41.4%	0.44%	49.1%	0.42%	49.9%
400000	1.25%	3.0%	0.41%	40.9%	0.50%	49.1%	0.48%	49.9%

Table B.3: Accuracy and sample size for the π estimator, accounting product B03

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.18%	52.1%	0.41%	50.3%	0.42%	50.0%	0.41%	50.0%
40000	0.18%	47.3%	0.40%	49.8%	0.45%	50.0%	0.45%	50.0%
60000	0.20%	42.6%	0.45%	49.3%	0.48%	49.9%	0.48%	50.0%
80000	0.31%	37.9%	0.38%	48.8%	0.39%	49.9%	0.39%	50.0%
100000	0.35%	33.6%	0.38%	48.3%	0.45%	49.8%	0.46%	50.0%
120000	0.43%	29.4%	0.39%	47.8%	0.43%	49.8%	0.44%	50.0%
140000	0.46%	25.6%	0.31%	47.3%	0.35%	49.7%	0.36%	50.0%
160000	0.50%	22.2%	0.28%	46.8%	0.37%	49.7%	0.37%	50.0%
180000	0.35%	19.1%	0.31%	46.3%	0.38%	49.6%	0.41%	50.0%
200000	0.33%	16.4%	0.41%	45.8%	0.44%	49.6%	0.47%	50.0%
220000	0.36%	14.0%	0.38%	45.3%	0.44%	49.5%	0.46%	49.9%
240000	0.48%	11.9%	0.44%	44.8%	0.51%	49.5%	0.52%	49.9%
260000	0.38%	10.1%	0.34%	44.3%	0.40%	49.4%	0.42%	49.9%
280000	0.43%	8.5%	0.31%	43.8%	0.41%	49.4%	0.41%	49.9%
300000	0.73%	7.2%	0.38%	43.3%	0.51%	49.3%	0.52%	49.9%
320000	1.00%	6.0%	0.38%	42.8%	0.51%	49.3%	0.50%	49.9%
340000	1.23%	5.1%	0.46%	42.3%	0.54%	49.2%	0.53%	49.9%
360000	1.12%	4.3%	0.44%	41.8%	0.54%	49.2%	0.53%	49.9%
380000	1.37%	3.6%	0.48%	41.4%	0.58%	49.1%	0.57%	49.9%
400000	1.41%	3.0%	0.54%	40.9%	0.67%	49.1%	0.64%	49.9%

Table B.4: Variability and sample size for the π estimator, accounting product B03

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.17%	50.6%	0.27%	50.1%	0.26%	50.0%	0.25%	50.0%
40000	0.28%	45.7%	0.22%	49.6%	0.27%	50.0%	0.27%	50.0%
60000	0.25%	40.9%	0.22%	49.1%	0.30%	49.9%	0.29%	50.0%
80000	0.27%	36.3%	0.19%	48.6%	0.29%	49.9%	0.28%	50.0%
100000	0.40%	31.9%	0.18%	48.1%	0.27%	49.8%	0.26%	50.0%
120000	0.46%	27.8%	0.21%	47.6%	0.27%	49.8%	0.25%	50.0%
140000	0.53%	24.0%	0.18%	47.1%	0.24%	49.7%	0.22%	50.0%
160000	0.51%	20.6%	0.23%	46.6%	0.25%	49.7%	0.24%	50.0%
180000	0.67%	17.6%	0.27%	46.1%	0.26%	49.6%	0.25%	50.0%
200000	0.73%	14.9%	0.28%	45.6%	0.27%	49.6%	0.26%	50.0%
220000	0.70%	12.6%	0.31%	45.1%	0.29%	49.6%	0.27%	50.0%
240000	0.64%	10.6%	0.33%	44.6%	0.33%	49.5%	0.31%	50.0%
260000	0.54%	8.8%	0.33%	44.1%	0.31%	49.5%	0.29%	50.0%
280000	0.43%	7.4%	0.37%	43.6%	0.34%	49.4%	0.31%	50.0%
300000	0.59%	6.1%	0.37%	43.1%	0.32%	49.3%	0.31%	50.0%
320000	0.63%	5.1%	0.35%	42.7%	0.31%	49.3%	0.29%	50.0%
340000	0.93%	4.2%	0.38%	42.2%	0.33%	49.3%	0.30%	50.0%
360000	0.81%	3.5%	0.39%	41.7%	0.37%	49.2%	0.35%	50.0%
380000	0.71%	2.9%	0.45%	41.2%	0.37%	49.2%	0.36%	50.0%
400000	0.96%	2.4%	0.46%	40.7%	0.35%	49.1%	0.34%	50.0%

Table B.5: Accuracy and sample size for the π estimator, accounting product B15

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.23%	50.6%	0.37%	50.1%	0.40%	50.0%	0.37%	50.0%
40000	0.37%	45.7%	0.30%	49.6%	0.35%	50.0%	0.32%	50.0%
60000	0.25%	40.9%	0.25%	49.1%	0.35%	49.9%	0.34%	50.0%
80000	0.30%	36.3%	0.25%	48.6%	0.35%	49.9%	0.34%	50.0%
100000	0.46%	31.9%	0.27%	48.1%	0.32%	49.8%	0.30%	50.0%
120000	0.56%	27.8%	0.28%	47.6%	0.31%	49.8%	0.29%	50.0%
140000	0.62%	24.0%	0.23%	47.1%	0.28%	49.7%	0.27%	50.0%
160000	0.64%	20.6%	0.27%	46.6%	0.30%	49.7%	0.29%	50.0%
180000	0.80%	17.6%	0.32%	46.1%	0.30%	49.6%	0.29%	50.0%
200000	0.91%	14.9%	0.33%	45.6%	0.31%	49.6%	0.30%	50.0%
220000	0.89%	12.6%	0.36%	45.1%	0.34%	49.6%	0.32%	50.0%
240000	0.82%	10.6%	0.39%	44.6%	0.38%	49.5%	0.36%	50.0%
260000	0.81%	8.8%	0.39%	44.1%	0.38%	49.5%	0.36%	50.0%
280000	0.64%	7.4%	0.43%	43.6%	0.40%	49.4%	0.38%	50.0%
300000	0.59%	6.1%	0.42%	43.1%	0.38%	49.3%	0.36%	50.0%
320000	0.73%	5.1%	0.39%	42.7%	0.36%	49.3%	0.34%	50.0%
340000	0.93%	4.2%	0.41%	42.2%	0.37%	49.3%	0.33%	50.0%
360000	0.72%	3.5%	0.43%	41.7%	0.40%	49.2%	0.37%	50.0%
380000	0.73%	2.9%	0.48%	41.2%	0.40%	49.2%	0.37%	50.0%
400000	1.11%	2.4%	0.51%	40.7%	0.41%	49.1%	0.37%	50.0%

Table B.6: Variability and sample size for the π estimator, accounting product B15

B.2 The regression estimator

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.06%	62.7%	0.11%	52.2%	0.16%	50.2%	0.15%	50.0%
40000	0.06%	58.6%	0.11%	51.8%	0.15%	50.2%	0.14%	50.0%
60000	0.03%	54.4%	0.11%	51.3%	0.12%	50.1%	0.13%	50.0%
80000	0.04%	50.2%	0.11%	50.8%	0.13%	50.1%	0.14%	50.0%
100000	0.04%	46.1%	0.14%	50.3%	0.12%	50.0%	0.13%	49.9%
120000	0.03%	42.2%	0.12%	49.8%	0.11%	50.0%	0.12%	49.9%
140000	0.04%	38.5%	0.14%	49.5%	0.12%	50.0%	0.14%	50.0%
160000	0.09%	34.9%	0.13%	49.0%	0.13%	50.0%	0.15%	50.1%
180000	0.06%	31.4%	0.13%	48.5%	0.14%	49.9%	0.16%	50.0%
200000	0.11%	28.5%	0.12%	48.0%	0.11%	49.9%	0.14%	50.0%
220000	0.13%	25.7%	0.14%	47.5%	0.10%	49.9%	0.12%	50.1%
240000	0.16%	23.2%	0.13%	47.0%	0.10%	49.8%	0.12%	50.1%
260000	0.21%	20.9%	0.13%	46.6%	0.11%	49.8%	0.13%	50.1%
280000	0.23%	18.8%	0.12%	46.1%	0.11%	49.7%	0.12%	50.1%
300000	0.24%	17.0%	0.09%	45.6%	0.11%	49.7%	0.12%	50.1%
320000	0.26%	15.4%	0.11%	45.2%	0.12%	49.7%	0.12%	50.2%
340000	0.32%	13.9%	0.08%	44.7%	0.11%	49.7%	0.11%	50.2%
360000	0.32%	12.6%	0.07%	44.2%	0.11%	49.6%	0.11%	50.2%
380000	0.35%	11.4%	0.07%	43.7%	0.11%	49.6%	0.11%	50.2%
400000	0.27%	10.4%	0.09%	43.2%	0.13%	49.5%	0.13%	50.1%

Table B.7: Accuracy and sample size for the regression estimator, accounting product B01

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.07%	62.7%	0.14%	52.2%	0.20%	50.2%	0.19%	50.0%
40000	0.07%	58.6%	0.13%	51.8%	0.19%	50.2%	0.18%	50.0%
60000	0.04%	54.4%	0.12%	51.3%	0.17%	50.1%	0.16%	50.0%
80000	0.04%	50.2%	0.13%	50.8%	0.17%	50.1%	0.18%	50.0%
100000	0.05%	46.1%	0.14%	50.3%	0.17%	50.0%	0.17%	49.9%
120000	0.04%	42.2%	0.15%	49.8%	0.16%	50.0%	0.16%	49.9%
140000	0.06%	38.5%	0.16%	49.5%	0.17%	50.0%	0.17%	50.0%
160000	0.12%	34.9%	0.16%	49.0%	0.17%	50.0%	0.17%	50.1%
180000	0.09%	31.4%	0.16%	48.5%	0.18%	49.9%	0.18%	50.0%
200000	0.13%	28.5%	0.15%	48.0%	0.15%	49.9%	0.16%	50.0%
220000	0.15%	25.7%	0.16%	47.5%	0.15%	49.9%	0.15%	50.1%
240000	0.15%	23.2%	0.14%	47.0%	0.15%	49.8%	0.14%	50.1%
260000	0.29%	20.9%	0.15%	46.6%	0.15%	49.8%	0.15%	50.1%
280000	0.32%	18.8%	0.15%	46.1%	0.16%	49.7%	0.15%	50.1%
300000	0.34%	17.0%	0.12%	45.6%	0.15%	49.7%	0.14%	50.1%
320000	0.37%	15.4%	0.14%	45.2%	0.17%	49.7%	0.16%	50.2%
340000	0.41%	13.9%	0.11%	44.7%	0.15%	49.7%	0.14%	50.2%
360000	0.41%	12.6%	0.10%	44.2%	0.16%	49.6%	0.14%	50.2%
380000	0.45%	11.4%	0.09%	43.7%	0.16%	49.6%	0.14%	50.2%
400000	0.32%	10.4%	0.12%	43.2%	0.18%	49.5%	0.17%	50.1%

Table B.8: Variability and sample size for the regression estimator, accounting product B01

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.04%	52.1%	0.05%	50.3%	0.06%	50.0%	0.05%	50.0%
40000	0.03%	47.3%	0.07%	49.8%	0.06%	50.0%	0.05%	50.0%
60000	0.03%	42.6%	0.05%	49.3%	0.06%	49.9%	0.05%	50.0%
80000	0.06%	37.9%	0.06%	48.8%	0.05%	49.9%	0.05%	50.0%
100000	0.06%	33.6%	0.06%	48.3%	0.04%	49.8%	0.04%	50.0%
120000	0.07%	29.4%	0.05%	47.8%	0.04%	49.8%	0.04%	50.0%
140000	0.08%	25.6%	0.03%	47.3%	0.04%	49.7%	0.04%	50.0%
160000	0.10%	22.2%	0.04%	46.8%	0.05%	49.7%	0.06%	50.0%
180000	0.10%	19.1%	0.04%	46.3%	0.06%	49.6%	0.06%	50.0%
200000	0.12%	16.4%	0.04%	45.8%	0.06%	49.6%	0.06%	50.0%
220000	0.12%	14.0%	0.04%	45.3%	0.06%	49.5%	0.06%	49.9%
240000	0.16%	11.9%	0.03%	44.8%	0.05%	49.5%	0.05%	49.9%
260000	0.12%	10.1%	0.05%	44.3%	0.06%	49.4%	0.07%	49.9%
280000	0.16%	8.5%	0.05%	43.8%	0.06%	49.4%	0.07%	49.9%
300000	0.16%	7.2%	0.06%	43.3%	0.06%	49.3%	0.07%	49.9%
320000	0.19%	6.0%	0.06%	42.8%	0.07%	49.3%	0.07%	49.9%
340000	0.25%	5.1%	0.06%	42.3%	0.06%	49.2%	0.07%	49.9%
360000	0.25%	4.3%	0.07%	41.8%	0.06%	49.2%	0.07%	49.9%
380000	0.30%	3.6%	0.06%	41.4%	0.05%	49.1%	0.06%	49.9%
400000	0.32%	3.0%	0.06%	40.9%	0.05%	49.1%	0.07%	49.9%

Table B.9: Accuracy and sample size for the regression estimator, accounting product B03

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.04%	52.1%	0.07%	50.3%	0.07%	50.0%	0.06%	50.0%
40000	0.04%	47.3%	0.08%	49.8%	0.07%	50.0%	0.07%	50.0%
60000	0.04%	42.6%	0.08%	49.3%	0.07%	49.9%	0.06%	50.0%
80000	0.07%	37.9%	0.07%	48.8%	0.06%	49.9%	0.06%	50.0%
100000	0.09%	33.6%	0.07%	48.3%	0.05%	49.8%	0.05%	50.0%
120000	0.09%	29.4%	0.06%	47.8%	0.05%	49.8%	0.05%	50.0%
140000	0.09%	25.6%	0.05%	47.3%	0.05%	49.7%	0.05%	50.0%
160000	0.12%	22.2%	0.05%	46.8%	0.06%	49.7%	0.07%	50.0%
180000	0.11%	19.1%	0.05%	46.3%	0.07%	49.6%	0.07%	50.0%
200000	0.15%	16.4%	0.05%	45.8%	0.07%	49.6%	0.08%	50.0%
220000	0.14%	14.0%	0.04%	45.3%	0.07%	49.5%	0.07%	49.9%
240000	0.18%	11.9%	0.04%	44.8%	0.05%	49.5%	0.05%	49.9%
260000	0.16%	10.1%	0.06%	44.3%	0.06%	49.4%	0.08%	49.9%
280000	0.21%	8.5%	0.05%	43.8%	0.06%	49.4%	0.07%	49.9%
300000	0.21%	7.2%	0.07%	43.3%	0.08%	49.3%	0.09%	49.9%
320000	0.25%	6.0%	0.08%	42.8%	0.07%	49.3%	0.08%	49.9%
340000	0.31%	5.1%	0.07%	42.3%	0.07%	49.2%	0.08%	49.9%
360000	0.31%	4.3%	0.08%	41.8%	0.06%	49.2%	0.08%	49.9%
380000	0.37%	3.6%	0.06%	41.4%	0.06%	49.1%	0.07%	49.9%
400000	0.42%	3.0%	0.06%	40.9%	0.06%	49.1%	0.08%	49.9%

Table B.10: Variability and sample size for the regression estimator, accounting product B03

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	relative error	sample size	relative error	sample size	relative error	sample size	relative error	sample size
20000	0.03%	50.6%	0.03%	50.1%	0.04%	50.0%	0.04%	50.0%
40000	0.02%	45.7%	0.04%	49.6%	0.03%	50.0%	0.03%	50.0%
60000	0.03%	40.9%	0.05%	49.1%	0.04%	49.9%	0.04%	50.0%
80000	0.03%	36.3%	0.05%	48.6%	0.05%	49.9%	0.05%	50.0%
100000	0.05%	31.9%	0.06%	48.1%	0.06%	49.8%	0.06%	50.0%
120000	0.07%	27.8%	0.05%	47.6%	0.05%	49.8%	0.05%	50.0%
140000	0.06%	24.0%	0.05%	47.1%	0.05%	49.7%	0.05%	50.0%
160000	0.06%	20.6%	0.05%	46.6%	0.05%	49.7%	0.04%	50.0%
180000	0.04%	17.6%	0.05%	46.1%	0.04%	49.6%	0.04%	50.0%
200000	0.05%	14.9%	0.05%	45.6%	0.05%	49.6%	0.04%	50.0%
220000	0.03%	12.6%	0.05%	45.1%	0.04%	49.6%	0.04%	50.0%
240000	0.06%	10.6%	0.05%	44.6%	0.05%	49.5%	0.05%	50.0%
260000	0.16%	8.8%	0.05%	44.1%	0.04%	49.5%	0.04%	50.0%
280000	0.16%	7.4%	0.04%	43.6%	0.04%	49.4%	0.04%	50.0%
300000	0.19%	6.1%	0.04%	43.1%	0.04%	49.3%	0.04%	50.0%
320000	0.20%	5.1%	0.04%	42.7%	0.04%	49.3%	0.04%	50.0%
340000	0.27%	4.2%	0.05%	42.2%	0.04%	49.3%	0.04%	50.0%
360000	0.30%	3.5%	0.04%	41.7%	0.04%	49.2%	0.04%	50.0.%
380000	0.33%	2.9%	0.04%	41.2%	0.04%	49.2%	0.04%	50.0%
400000	0.31%	2.4%	0.04%	40.7%	0.04%	49.1%	0.04%	50.0%

Table B.11: Accuracy and sample size for the regression estimator, accounting product B15

	$k = 10^{-5}$		$k = 10^{-6}$		$k = 10^{-7}$		$k = 10^{-8}$	
x_0	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size	standard deviation	sample size
20000	0.03%	50.6%	0.04%	50.1%	0.05%	50.0%	0.05%	50.0%
40000	0.03%	45.7%	0.05%	49.6%	0.04%	50.0%	0.04%	50.0%
60000	0.04%	40.9%	0.07%	49.1%	0.06%	49.9%	0.06%	50.0%
80000	0.04%	36.3%	0.07%	48.6%	0.07%	49.9%	0.07%	50.0%
100000	0.06%	31.9%	0.07%	48.1%	0.07%	49.8%	0.07%	50.0%
120000	0.08%	27.8%	0.06%	47.6%	0.06%	49.8%	0.06%	50.0%
140000	0.06%	24.0%	0.05%	47.1%	0.05%	49.7%	0.05%	50.0%
160000	0.07%	20.6%	0.06%	46.6%	0.05%	49.7%	0.05%	50.0%
180000	0.06%	17.6%	0.06%	46.1%	0.05%	49.6%	0.05%	50.0%
200000	0.06%	14.9%	0.06%	45.6%	0.05%	49.6%	0.05%	50.0%
220000	0.05%	12.6%	0.06%	45.1%	0.05%	49.6%	0.05%	50.0%
240000	0.07%	10.6%	0.06%	44.6%	0.05%	49.5%	0.05%	50.0%
260000	0.18%	8.8%	0.05%	44.1%	0.05%	49.5%	0.05%	50.0%
280000	0.19%	7.4%	0.05%	43.6%	0.04%	49.4%	0.04%	50.0%
300000	0.22%	6.1%	0.05%	43.1%	0.04%	49.3%	0.04%	50.0%
320000	0.25%	5.1%	0.06%	42.7%	0.05%	49.3%	0.05%	50.0%
340000	0.30%	4.2%	0.06%	42.2%	0.05%	49.3%	0.05%	50.0%
360000	0.39%	3.5%	0.05%	41.7%	0.05%	49.2%	0.05%	50.0%
380000	0.43%	2.9%	0.05%	41.2%	0.05%	49.2%	0.05%	50.0%
400000	0.43%	2.4%	0.05%	40.7%	0.05%	49.1%	0.05%	50.0%

Table B.12: Variability and sample size for the regression estimator, accounting product B15

Bibliography

[1] G Andersson, *Livförsäkringsmatematik, andra upplagan*, Svenska Försäkringsföreningen 2013

[2] Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)

[3] Försäkringsrörelselagen, Svensk författningssamling 2010:2043

[4] M Koller, *Stochastic Models in Life Insurance*, Springer Science & Business Media 2012

[5] T Møller, M Steffensen, *Market-Valuation Methods in Life and Pension Insurance*, Cambridge University Press 2007

[6] QIS5 Technical Specifications, Annex to Call for Advice from CEIOPS on QIS5, CEIOPS 2010

[7] S M Ross, *Introduction to Probability Models, Eleventh Edition*, Elsevier Inc. 2014

[8] C-E Särndal, B Swensson, J Wretman, *Model Assisted Survey Sampling*, Springer-Verlag New York, Inc. 1992

[9] B von Bahr, G Ronge, *Trafikljus – stresstest för försäkrings- och driftskostnadsrisker inom livförsäkring*, Finansinspektionen 2007