

# Scenario Generation for Asset and Liability Management Models Applied to a Saudi Arabian Pension Fund

Maram Alwohaibi<sup>1,2,3</sup>, Diana Roman<sup>3\*</sup>, Alina Peluso<sup>3</sup>

<sup>1</sup>Department of Mathematics, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>2</sup>Basic and Applied Scientific Research Center, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>3</sup>Department of Mathematics, College of Engineering, Design & Physical Sciences, Brunel University London, Uxbridge, UK

Email: \*diana.roman@brunel.ac.uk

**How to cite this paper:** Alwohaibi, M., Roman, D., & Peluso, A. (2022). Scenario Generation for Asset and Liability Management Models Applied to a Saudi Arabian Pension Fund. *Journal of Financial Risk Management*, 11, 277-295.  
<https://doi.org/10.4236/jfrm.2022.112014>

**Received:** December 2, 2021

**Accepted:** May 8, 2022

**Published:** May 11, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In Asset and Liability Management (ALM) models, there are parameters whose values are not known with certainty at decision time, such as future asset returns, liability and contribution values. Simulation models generate possible “scenarios” for these parameters, which are used as inputs in the optimisation models and help thus in making decisions. These decisions can be evaluated in the sample, on the same scenarios that were used for making the decision, and out-of-sample, on a different, usually much larger, scenario set. With asset return simulation, the major difficulty lies in the multivariate nature of the data. We propose to capture this via the historical copula, making thus no distributional assumptions. We suggest the use of univariate sample generation which allows for different asset returns to be modelled by different distributions. The liabilities and contributions values have as a main source of uncertainty the population numbers; we propose to model this by adapting a model used in biology (BIDE). We use the resulting scenario generator in four different ALM optimisation models, using a dataset from the largest Saudi Arabian pension fund and the Saudi Arabian market index.

## Keywords

Asset and Liability Management, Liability Driven Investment, Risk Management, Funding Ratio, Population Modelling, Historical Copula

---

## 1. Introduction and Motivation

Market fluctuation and population aging require pension funds to adopt a new view on their asset allocation decisions. Asset and Liability Management (ALM) models have become well-established decision tools for pension funds. Other financial organisations such as banks, insurance companies, and even some wealthy individuals are also underpinned by the balancing of cash flow to match and outperform some future obligations or liabilities. Asset management techniques that take into account the stochastic nature of liabilities are given the generic label: Asset and Liability Management techniques (ALM) which have been more recently renamed by some authors as Liability Driven Investment (LDI) [Schwaiger \(2009\)](#).

While pure asset allocation problems are usually modeled as a single period, in ALM, the presence of liabilities to be paid over many future periods raises the need to adopt a multi-period setting. Thus, ALMs are commonly modelled as multi-stage optimisation models, in which a large terminal wealth is required, while at intermediate periods, constraints on the funding ratio, that is, the ratio of assets to liabilities, are imposed; see for example [Alwohaibi & Roman \(2018\)](#).

In ALM models, the outcome of investment decisions depends on the future realisations of parameters that are not known with certainty at decision time. For pension funds, such parameters are future asset returns, liabilities (money to be paid out), and contributions (money paid in). These parameters are described by distributions, if we consider one planning period, or stochastic processes, in the case of a multi-period planning horizon. In order to use these parameters in optimisation models, one needs to approximate these distributions as discrete, described by a representative set of “scenario” outcomes with corresponding probabilities, other said, to “generate scenarios”. Simulation of the unknown parameters is necessary not only for making decisions as above, but also for evaluating the decisions. Out-of-sample evaluation means evaluating a decision using a different set of scenarios than the one employed for making the decision, usually much larger. A special type of out of sample evaluation is stress testing, in which particularly unfavorable samples are being used for the purpose of testing performance under worst-case situations.

In this paper, we simulate the parameters of ALM models (liabilities, contributions, and asset returns) as follows.

The liabilities and contributions have the same main underlying source of uncertainty, namely the population numbers; once the numbers of members (paying contributors) and past members (retirees to whom liabilities are to be paid) are simulated, the corresponding contributions and liability values can be obtained using salary models and rules for pension payments. The plan’s demographic dynamics could be analysed either in a closed system or in an open system (the latter allowing for new members to be considered). Markov processes have been commonly used to describe the population dynamics; see for example [Mettler \(2005\)](#) which describes population dynamics in a closed and open sys-

tem for a defined benefit (DB) pension plan.

We propose to model population numbers by adapting a model based on the “Birth, Immigration, Death and Emigration” (BIDE) concept, originally used in biology [Nathan \(2016\)](#). The motivation for this is the fact that the dynamics over time in pension funds populations are similar; in addition, it is a more intuitive and easier model to implement.

Scenario generation methods for asset prices, or asset returns, have been extensively researched, mainly in the context of single period asset allocation. For an overview of scenario generation methods applied in finance and economic decision-making, see [Vázsonyi \(2006\)](#). Commonly used methods include sampling or bootstrapping of historical data [Efron \(1979\)](#) and the Vector Autoregression model (VAR) [Sims \(1980\)](#). Scenario generation using VAR in the area of ALM has been used in [Dert \(1995\)](#), [Kouwenberg \(2001\)](#), and [Sheikh Hussin \(2012\)](#).

Another established scenario generating method is the moment matching approach [Høyland & Wallace \(2001\)](#); it has been used in financial applications, including ALM, see for example [Dupačová & Polívka \(2009\)](#), [Fleten et al. \(2002\)](#) and [Kouwenberg \(2001\)](#). In this approach, the decision-maker specifies a set of statistical properties (e.g. moments of order up to four). The scenario set is constructed in such a way that these statistical properties are matched.

A major difficulty with scenario generation for asset prices is the multi-variate nature of data. One way to overcome this is to separately model the univariate marginal distributions and the dependencies between random variables via a copula [Sklar \(1959\)](#). This separation allows to combining approaches that capture shape with one method and handle margins with another. It allows to taking advantage of the versatility of statistical software packages (such as R, used in this work), which can fit a set of samples to a univariate distribution and generate further samples from the fitted univariate distribution. Different univariate distributions can be fitted to the marginals, rather than assuming that all asset returns follow the same distribution.

Various copulas can be used in order to satisfy specific assumptions on data dependency. Following [Kaut & Wallace \(2011\)](#), we use an *empirical copula*, thus making no parametric assumptions but only modelling dependency as shown by historical data. Different scenario sets are obtained by sampling again from the univariate distributions and combining the univariate sets of samples via the empirical copula.

To our knowledge, BIDE models have not been adapted and applied so far in the context of ALM models. Another contribution is the application of historical copula together with univariate sampling as a scenario generator of asset prices in ALM models. The two scenario generators, modelling different sources of randomness (population and asset prices) are combined in order to generate parameters in multi-period optimisation models of liability-driven investment.

The rest of the paper is organised as follows. In Section 2 we describe the sce-

nario generation method for asset returns, based on empirical copula and univariate sample generation. Section 3 concerns the BIDE model and its application to ALM, in order to generate scenarios for contribution and liability values. In Section 4 we shortly describe the optimisation models, for which the parameters (asset returns, liability, and contribution values) were obtained with the scenario generation methods in Sections 2 and 3. Numerical results are presented in Section 5, using a dataset drawn from GOSI, the largest Saudi Arabian pension fund, and TASI, the Saudi Arabian stock market. Conclusions are presented in Section 6.

## 2. Simulation of Multivariate Distributions Using Copula

In many applications, it is required to generate samples from multivariate distributions. While here we consider the specific case of simulating future asset returns, this approach is suitable for any type of data with a sufficient number of past observations available.

There are well-established methods for fitting samples into univariate distributions, implemented into standard statistical software, such as R (Rigby & Stasinopoulos, 2005; Stasinopoulos & Rigby, 2007). Not the same is available for multivariate distributions, where the dependency between random variables needs to be taken into account. Here, we propose to model this dependency via empirical copulas, similar to Kaut & Wallace (2011).

The name copula was first used by Sklar (1959) to define a tool that describes the multi-variate structure of a distribution (the dependence between the variables) irrespective of the marginal distributions. Using copulas allows separating the multivariate structure from the marginal distributions, thus allowing the marginals to be independently modelled. It also overcomes some limitations of other methods, such as using a correlation matrix.

A  $d$ -dimensional copula is the joint cumulative distribution function (cdf) of any  $d$ -dimensional random vector with standard uniform marginal distributions, *i.e.* a function  $C: [0,1]^d \rightarrow [0,1]$  Sklar (1959). Sklar's theorem (Sklar, 1959) states that any multivariate distribution can be written in terms of univariate marginal distribution functions and a copula:

Let  $F$  be a  $d$ -dimensional joint cumulative distribution function of a random vector  $(X_1, X_2, \dots, X_d)$  with margins  $F_1, \dots, F_d$ ,  
 $F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d)$ . Then, there exists a  $d$ -dimensional copula  $C$  such that for all  $x$  in  $\mathbb{R}^d$ .

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

Moreover, if all the marginal cdfs  $F_i$  are continuous, then  $C$  is unique Nelsen (2007).

An immediate consequence is that, for every  $(u_1, \dots, u_d) \in [0,1]^d$ ,

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

where  $F_i^{-1}$  is the generalised inverse of  $F_i$ :

$$F_i^{-1}(u) = \inf \{t : F_i(t) \geq u\}, \quad u \in [0,1]$$

Just like distributions, copulas have many parametric families with specialized methods for generation (see for example Nelsen (2007)). In this paper, we use a special kind of copula, the so-called *empirical copula*. We generate samples for each univariate margin; using the copula, the univariate samples are combined to form a sample from the multivariate distribution.

The basic motivation for using this method is to generate, starting from a large (historical) sample of multivariate data, other sets of samples of the same size.

Suppose we have available  $N$  samples (observations) from a  $d$ -variate distribution of a random vector  $(X_1, X_2, \dots, X_d)$ ; denote these samples by

$\mathcal{S} = \{(x_1^i, x_2^i, \dots, x_d^i)\}_{i=1}^N$ ; our goal is to generate a matrix  $X$  of the size  $N \times d$  of other samples from  $(X_1, X_2, \dots, X_d)$  using the empirical copula.

The main idea is to create a matrix  $N \times d$  of “ranks”; element  $c_j^i$  in this matrix is  $\frac{k}{N}$ , where  $k$  is the “rank” of observation  $x_j^i$  among the observed values of variable  $X_j$ . That is, element  $c_j^i$  corresponds to the  $k$ -th worst value out of the values of the observations  $x_j^i$ ,  $i = 1, \dots, N$  of the random variable  $X_j$ ,  $j = 1, \dots, d$ . In this approach, instead of making distribution assumptions on the marginal distribution functions  $F_j$ ,  $j = 1, \dots, d$ , we use instead the empirical distributions with marginal cdfs given by:

$$F_j^e(x_j^i) = \frac{\text{rank}(x_j^i, x_j)}{N}, \quad j = 1, \dots, d; \quad i = 1, \dots, N$$

where  $\text{rank}(x^i, x)$  is the rank (order) of value  $x^i$  in a vector  $x$ , with values between 1 and  $N$ .

Thus, we can interpret a row of this matrix as a “scenario” of dependence between the  $d$  random variables; for example, one scenario may be the maximum of margin 1 occurs at the same time with the second-worst value of margin 2 together with the minimum of margin 3, etc.

Once  $N$  samples from each of the univariate distributions are (independently) generated, they can be combined according to the matrix of ranks in order to form a new scenario set of the multivariate data.

### 3. BIDE Population Models Applied to Pension Fund Population

The scenarios for the contributions and liabilities in an ALM model have the same main underlying source of uncertainty, namely the population numbers.

We adopt here a population model used in biology, in order to model over time both the population of contributors and that of retirees in a pension fund.

The dynamics of a population in a BIDE model (Birth, Immigration, Death, Emigration model) Nathan (2016) is given by:

$$N_{t+1} = N_t + B_t - D_t + I_t - E_t$$

where:

- $N_t$  represents the population size at time  $t$
- $B_t$  is the number of births within the population between  $N_t$  and  $N_{t+1}$ ;
- $D_t$  is the number of deaths within the population between  $N_t$  and  $N_{t+1}$ ;
- $I_t$  is the number of individuals immigrating into the population between  $N_t$  and  $N_{t+1}$ ;
- $E_t$  is the number of individuals emigrating from the population between  $N_t$  and  $N_{t+1}$ .

In the remaining of this section, we explain how we construct sample paths for these populations and how to generate the scenarios for the contributions and liability payments.

### The Contributors' and Retirees' Population Sample Paths

The BIDE equation is adapted as follows:

$$N_{t+1,s} = N_{t,s} + New_{t,s} - R_{t,s} + TI_{t,s} - TO_{t,s} - D_{t,s}, \quad t = 0, 1, \dots, T-1; s = 1, \dots, S \quad (1)$$

where  $T$  is the number of time periods considered (e.g. years),  $S$  is the number of sample paths to generate and:

- $N_{t,s}$  the total number of contributors in employment by the end of time period  $t$  under scenario  $s$
- $New_{t,s}$  the numbers of new employees between time  $t$  and  $t+1$  under scenario  $s$
- $R_{t,s}$  the total number of contributors who leave the scheme (due to retirement or death) between  $t$  and  $t+1$  under scenario  $s$
- $TI_{t,s}$  the total number of employees who enter the system from another pension fund (*i.e.* transferred) between  $t$  and  $t+1$  under scenario  $s$
- $TO_{t,s}$  the total number of employees who leave the system and are transferred to another pension fund between time  $t$  and  $t+1$  under scenario  $s$
- $D_{t,s}$  the total number of cases that received lump sum payments and leave the scheme between times  $t$  and  $t+1$  under scenario  $s$ .

If past observations are available, we can compute a set of “observed” ratios as follows. Let  $\gamma_t$  denote the employment ratio at time  $t$ , defined as:

$$\gamma_t = New_t / N_t$$

Similarly,  $\mu_t$  is the ratio of the retirement (including deaths) at time  $t$ , defined as:

$$\mu_t = R_t / N_t$$

Similarly, let us denote by  $\eta_t$  the ratio of the transfers out of the fund at time  $t$ :

$$\eta_t = TI_t / N_t$$

$\phi_t$  is the ratio of the transfers out of the fund at time  $t$

$$\phi_t = TO_t / N_t$$

Finally,  $\Delta_t$  represent the ratio of leaving the scheme at time  $t$  defined as:

$$\Delta_t = D_t / N_t$$

Each of these ratios at the current time is a random variable that affects the total number of contributors the next time period. By sampling from these observed (historical) ratios we obtain a vector  $(\gamma, \mu, \eta, \phi, \Delta)$  that represents a possible scenario of these ratios for the next time period. Using the current (known) number of contributors  $N_t$  and the scenario for ratios, we can simulate the number of new contributors and so forth until the next time period:

$$New_t = \gamma_t N_t; \quad R_t = \mu_t N_t; \quad TI_t = \eta_t N_t; \quad TO_t = \phi_t N_t; \quad D_t = \Delta_t N_t$$

Using (1) we obtain a simulated value for the number of contributors at the next time period  $N_{t+1}$ . Repeating the process, using  $N_{t+1}$  and another sample of ratios  $(\gamma, \mu, \eta, \phi, \Delta)$  we obtain a scenario value for  $N_{t+2}$ , and so forth. We can thus construct a sample path for the contributors' population.

Similarly, we adopt the BIDE model to represent the dynamics of the retirees population:

$$NR_{t+1,s} = NR_{t,s} + R_{t,s} - G_{t,s}, \quad t = 0, 1, \dots, T-1, \quad s = 1, \dots, S \quad (2)$$

where:

- $NR_{t,s}$  the total number of retirees who receive pension at time  $t$  under scenario  $s$  (*i.e.* retirees' population size at time  $t$  under scenario  $s$ );
- $G_{t,s}$  the number of cases that stop receiving pension (leaving the retirees population) between times  $t$  and  $t+1$  under scenario  $s$ ;
- $R_{t,s}$  the total number of contributors who leave the scheme and enter the retirees' population between times  $t$  and  $t+1$  under scenario  $s$ .

A similar approach, based on observed rates, that is, percentages of cases leaving/entering the population out of the initial population, is employed in order to construct sample paths for the retirees' population numbers.

Sample paths for the cash inflows (contributions) and outflows (liabilities) can be generated using salary models and specific payment rules of the considered funds. In Section 5 we do this, for a pension fund in Saudi Arabia.

#### 4. ALM Optimisation Models

We consider ALM models with an initial portfolio of financial assets; decisions are to be made regarding the rebalancing of this portfolio at specific times such that, a long-term wealth growth is achieved, while liabilities are satisfied at all times. It is thus common to have the terminal wealth fund as an objective function (to maximise) while intermediate-risk constraints are imposed. These risk constraints refer to the funding ratio (that is, the ratio of assets value to liabilities value) being kept acceptably high, at all intermediate time points. Usually, a target value for the funding ratio (such as 1.1 or 1.2) is specified as a parameter of the model.

Some of the most established models in this category include the integrated chance constraint model (ICCP) (Klein Haneveld et al., 2010) and the Maximin model (Young, 1998), which can be viewed as a particular case of the CVAR-ALM

model of Uryasev et al. (Bogentoft et al., 2001). With the ICCP model, a risk constraint is imposed, requiring that the expected value of the shortfall of the funding ratio with respect to the target is no more than a specified limit. Maximin models find the solution whose corresponding funding ratio performs the best under the worst-case scenario.

The formulations of the models are given in the Appendix.

A different approach is suggested in Alwohaibi & Roman (2018), where the risk of underfunding is modelled based on the concept of stochastic dominance Whitmore & Findlay (1978). Here, investment decisions are taken such that the distribution of the funding ratio is non-dominated with respect to Second-order Stochastic Dominance (SSD) and also is close in an optimal sense to a user-specified target distribution. The terminal wealth is specified in a constraint, rather than as an objective.

Two SSD-based models are developed, SSD-scaled and SSD-unscaled, depending on whether we compare scaled tails or unscaled tails of the distribution of funding ratio with the corresponding tails of the target distribution. Scaled tails are, roughly speaking, averages of worst-case outcomes while unscaled tails are sums of worst-case outcomes. It is developed in Alwohaibi & Roman (2018) that the ICCP and the Maximin are at two opposite extremes, one looking at the expected value of the shortfall (of funding ratio with respect to target) and the other looking at the worst-case scenario. In between, there is a multitude of solutions offering different shapes of the funding ratio distributions, with different trade-offs between a good left tail and a low value of expected shortfall. As pointed out in Alwohaibi & Roman (2018), the SSD scaled solutions are generally closer to Maximin solutions, in the sense that they offer good left tails of the funding ratio distributions; the SSD unscaled solutions are closer to the ICCP solutions, in the sense of having a low value of expected shortfall.

The algebraic formulations of the four optimisation models (ICCP, Maximin, SSD scaled, and SSD unscaled), together with parameter choices for the target funding ratio (in ICCP), distribution of the target funding ratio, and the minimum terminal wealth (in the SSD models) are given in the Appendix.

## 5. Numerical Results

### 5.1. Dataset and Computational Setup

We use the simulation models described in Sections 2 and 3 in order to generate scenarios for asset returns and liability and contribution values in ALM optimisation models, applied to a Saudi Arabian pension fund. We use the four optimisation models described in Section 4 and formulated in the Appendix; they are implemented in AMPL Fourer et al. (1990) and solved using CPLEX 12.5.1.0. The decisions obtained are then evaluated out-of-sample, using much larger scenario sets generated with the same methods.

As in Consigli & Dempster (1998) and Mulvey et al. (2000), the planning horizon is 10 years;  $t = 0$  refers to the year 2016. We consider for investment 16

asset classes: the Saudi equities represented by 15 sectors indices and also cash. Investment decisions have to be taken “now” ( $t = 0$ ) and then rebalanced every year,  $t = 1, \dots, 9$ .

For generating the scenarios for liabilities and contributions, we consider a dataset drawn from the General Organisation for Social Insurance (GOSI) [The General Organization for Social Insurance \(GOSI\) \(n.d.\)](http://www.gosi.gov.sa/) website: <http://www.gosi.gov.sa/>. The dynamics of the pension fund population are described by a BIDE-type model, as per section 3. We use historical data from the GOSI’s population as an input; that includes for example the number of participants, number of retirees, employment, and retirement rates for the last 10 years. It also includes the average salary each year and average salary growth. We use a simple salary model assuming a constant growth rate each year. We follow the GOSI-specific regulations in setting the percentage of salary to be paid in, as contributions, or out, as liabilities. For more details, please see [Alwohaibi & Roman \(2018\)](#).

The in-sample scenarios for the asset returns are obtained by bootstrapping from historical data drawn from the Saudi Arabian stock market index (TASI) [Saudi Stock Exchange \(n.d.\)](https://www.tadawul.com.sa) website: <https://www.tadawul.com.sa> for the period Jun 2007 to Nov 2015; we bootstrap 30 yearly rates of return. For the risk-free rate of return (interest rate) we consider the current Saudi Arabian interest rate of 2% following [Trading Economics \(n.d.\)](http://www.tradingeconomics.com/saudi-arabia/interest-rate) website: <http://www.tradingeconomics.com/saudi-arabia/interest-rate>.

The scenarios for asset returns are combined with 10 scenarios for contributions and liabilities, resulting thus in an in-sample scenario set of 300 sample paths.

The out-of-sample analysis is conducted over 11 different, larger, data sets. As optimisation is not employed, much larger datasets can be used.

For the asset returns, the first set is obtained by considering all the observed historical returns of the component assets; we compute 1937 scenarios for the annual rates of returns of the assets. The rest of the data sets for the asset returns are of the same size and are created by employing a historical copula, as described in Section 2, and sampling from the marginals. For each marginal, we use the historical samples and fit them into a univariate distribution, using the **R** package *gamlss* ([Rigby & Stasinopoulos, 2005](#)); different distributions are obtained. For the sake of brevity, we do not include the full details here; these can be obtained upon request. We then generate other 1937 samples from the fitted distribution for each margin and combine them via the empirical copula and repeat this process to create other 10 out-of-sample scenario sets for asset returns, in addition to the “historical” set. Each of these sets is then combined with 500 scenarios for liabilities generated in the same manner as the in-sample data sets.

To summarise, the out-of-sample analysis is conducted over 11 sets of scenarios each of size 968,500.

The out-of-sample analysis is summarised below:

- 1) Generate the in-sample scenarios for the optimisation problems.
- 2) Solve the models (SSD-Unscaled), (SSD-Scaled), (ICCP), and (Maximin) using the in-sample scenarios.
- 3) Generate 11 larger sets of out-of-sample scenarios.
- 4) Use the first stage investment decisions obtained at 2 and compute the realisations of the rate of returns distribution and the funding ratio distribution, considering an out-of-sample scenario set generated in 3.
- 5) Compute performance and risk-adjusted performance measures.
- 6) Repeat the last two steps for each of the 11 out-of-sample scenario sets.

## 5.2. Computational Results

We are interested in the funding ratio distributions (that is, the ratio of asset value to liabilities) at the first stage, in the sample, and out-of-sample. We compute key statistics from these distributions and investigate whether the out-of-sample statistics indicate the same key characteristics of the funding ratios as indicated in-sample.

**Table 1** presents in-sample results regarding the funding ratio distributions in each of the four optimisation models considered. Apart from the expected value, we are interested in the expected value of shortfall with respect to a target funding ratio of 1.10 and in the left tails. The  $A\%$  tail is defined as the average of the worst  $A\%$  of the outcomes. We thus consider not only the worst-case scenarios in each of the distributions (the minimum) but also a progressively higher number of worst-case scenarios. For all of the above statistics apart from the expected shortfall, high values are desirable.

The in-sample statistics are in line with the stated purposes of the optimisation models. That is, the Minimax distribution has the best (highest) worst-case value at 0.8932; moreover, it has the best-left tails when considering up to 25% of the worst-case scenarios: 0.9551 for the 25% Scaled tail. It also has, on the other

**Table 1.** Statistics of the in-sample funding ratios: (SSD-Unscaled), (SSD-Scaled), (ICCP) and (Maximin) models.

Statistic	SSD-Unscaled	SSD-Scaled	ICCP	Maximin
Expected value	1.160	1.154	1.141	1.135
Minimum	0.8168	0.876	0.7926	0.8932
Expected shortfall w.r.t 1.10	0.0455	0.0526	0.0430	0.0517
1%-Scaled tail	0.8200	0.8794	0.7957	0.8967
5%-Scaled tail	0.8526	0.8969	0.8283	0.9124
10%-Scaled tail	0.8725	0.9165	0.8556	0.9290
15%-Scaled tail	0.8946	0.9284	0.8915	0.9392
20%-Scaled tail	0.9207	0.9383	0.9227	0.9471
25%-Scaled tail	0.9420	0.9477	0.9464	0.9551

hand, the lowest expected value at 1.135.

The ICCP distribution has the lowest expected shortfall at 0.0430; it also has, however, the lowest minimum at 0.7926 and the worst left tails when considering up to 15% of the lowest outcomes.

The SSD distributions are somewhat in between; their expected values are the highest, while at the same time, the left tails are reasonably high. The SSD scaled distribution is closer to the Minimax one, in the sense of having good left tails while the SSD unscaled distribution is close to the ICCP one, in the sense of having a low expected shortfall (also the highest mean). It is developed in [Alwohaibi & Roman \(2018\)](#) that the Minimax and ICCP models can be obtained as particular cases of the SSD scaled and SSD unscaled models, respectively.

The first stage solution is used together with the 11 out-of-sample scenario sets (described in Section 5.1) in order to compute 11 out-of-sample distributions of funding ratios.

**Table 2** displays the expected values of the 11 out-of-sample distributions of funding ratios. Encouragingly, although the values are marginally lower, they are very similar to those obtained in the sample, although the number of out-of-sample scenarios is much higher. Importantly, the out-of-sample values preserve the ranking within the four models, in the sense of the highest means being those of the SSD-based distributions (on average, 1.1184 and 1.1126) and the lowest means those of the Minimax distribution (on average, 1.0982).

In addition to expected values, we look at the left tails of the out-of-sample funding ratio distributions; we compute the 1%, 5%, 10%, 15%, 20%, and 25% scaled left tails of each of the out of sample distributions. For the sake of brevity,

**Table 2.** The expected values of the out of sample funding ratio distributions, obtained using 11 out-of-sample scenario sets.

Dataset	SSD-Unscaled	SSD-Scaled	ICCP	Maximin
Data set 1	1.1198	1.1169	1.1036	1.1020
Data set 2	1.1198	1.1148	1.1052	1.1002
Data set 3	1.1227	1.1177	1.1087	1.1026
Data set 4	1.1225	1.1108	1.1083	1.0963
Data set 5	1.1171	1.1133	1.1026	1.0989
Data set 6	1.1193	1.1156	1.1054	1.1009
Data set 7	1.1195	1.1133	1.1060	1.0988
Data set 8	1.1139	1.1140	1.1006	1.0995
Data set 9	1.1175	1.1121	1.1029	1.0977
Data set 10	1.1153	1.1073	1.1025	1.0934
Data set 11	1.1178	1.1097	1.1034	1.0956
Average	1.1184	1.1126	1.1045	1.0982
St. Deviation	0.0030	0.0031	0.0028	0.0028

we include here the results for the 5% tails, as they are representative; the full set of results can be obtained upon request.

**Table 3** presents the out-of-sample results concerning the 5% scaled tails of the funding ratio. While these values are lower than the corresponding in-sample ones, this is to be expected, as the out-of-sample scenario sets are much bigger, accounting for more unfavorable outcomes.

Very importantly, the out-of-sample scenario sets distributions preserve the same pattern as the in-sample ones, corresponding to the stated purposes of the optimisation models. The “best” left tail corresponds to the Maximin distribution, followed by the SSD scaled one. As shown in **Table 3**, the average of the 5% worst-case outcomes is around 0.8161 in the case of Maximin distribution, considerably higher than the case of ICCP, which on average is at 0.7201; the “worst” left tail is that of the ICCP distribution, as computed on all out of sample datasets.

## 6. Conclusion and Further Research

We have proposed scenario generation models for the uncertain parameters of ALM models based on:

- 1) the BIDE model for simulating population numbers;
- 2) historical copula and univariate sample generation for simulating multivariate distributions such as asset returns.

The BIDE model has been used in biology but it can be adapted to populations describing both contributors and retirees in an ALM. Combined with salary models and the specific rules of payment and contribution of a fund, it can be used to generate scenarios for liability values and contribution values, which are ultimately the parameters required in ALM models.

**Table 3.** The 5% scaled tails of the out of sample funding ratio distributions for the 11 out-of-sample data sets.

Sample number	SSD-Unscaled	SSD-Scaled	ICCP	Maximin
Data set 1	0.7089	0.7810	0.6964	0.8103
Data set 2	0.7273	0.7904	0.7221	0.8182
Data set 3	0.735	0.7927	0.7298	0.82
Data set 4	0.7344	0.7914	0.7288	0.8186
Data set 5	0.7217	0.7865	0.7193	0.815
Data set 6	0.7253	0.7872	0.721	0.8156
Data set 7	0.7263	0.7879	0.7202	0.8159
Data set 8	0.7251	0.7889	0.7205	0.8169
Data set 9	0.7273	0.7893	0.7226	0.8171
Data set 10	0.7215	0.7834	0.7153	0.8121
Data set 11	0.7301	0.7905	0.725	0.8179
Average	0.7257	0.7881	0.7201	0.8161
St. Deviation	0.0071	0.0035	0.0089	0.0029

For asset returns, there is usually a large number of historical observations available; the use of historical copula is based on the assumption that the history captured well the dependency between asset returns. The purpose is to generate more scenario sets than available via historical data, including ideally those that would capture unfavorable scenarios. In our computational work, we used R in order to fit univariate distributions (individual asset returns) and generate further samples. This offers good flexibility, as very different univariate distributions may fit. The samples generated from univariate distributions are then combined together via the historical copula.

We have used the resulting scenario generators in four ALM optimisation models. The first stage solutions obtained, representing investment decisions, are evaluated in-sample (on the same scenario set, of size 300, used in order to obtain them) and out-of-sample (on several, much larger, scenario sets, not used in the optimisation process). More precisely, we looked at the distribution, in sample and out-of-sample, of the funding ratio: the ratio of assets value to liabilities. We considered expected values and “left tails”: averages of progressively higher percentages of worst-case scenarios. The results are encouraging: the out-of-sample distributions have very similar expected values to the in-sample ones. While the left tails are worse in the out-of-sample cases, (which is expected, as more unfavorable outcomes are taken into consideration), the shape of the out-of-sample distributions is similar to the in-sample case, in accordance with the stated purposes of the optimisation models. For example, the solution of the Minimax model, which in-sample has the highest worst-case value of the funding ratio (and generally best-left tails) with the lowest expected value of funding ratio out of the four models considered, has the same characteristics evaluated out of sample. This is a very important aspect, as it shows that the scenario generators work well with the optimisation models for ALM.

In our research, when applying the BIDE model in order to simulate future population numbers, the number of newcomers (“births”), retirees, etc. are obtained considering past (observed) proportions out of the total population. This is an important limitation; one way to overcome this is by considering additional scenarios, particularly those accounting for more recent trends such as the higher lifetime of contributing members and higher number of years in retirement.

We also considered a simple salary model, assuming a constant increase rate; more realistic salary models may offer a better picture. Finally, our scenario tree is a fan one: after generating scenarios for the first period, each node results in a single scenario for the next period. While the usual multi-period scenario tree would have obvious implications e.g. in terms of computational time, it would also provide a more realistic representation.

### **Conflicts of Interest**

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Alwohaibi, M., & Roman, D. (2018). ALM Models Based on Second Order Stochastic Dominance. *Computational Management Science*, *15*, 187-211. <https://doi.org/10.1007/s10287-018-0299-8>
- Bogentoft, E., Edwin Romeijn, H., & Uryasev, S. (2001). Asset/Liability Management for Pension Funds Using CVaR Constraints. *Journal of Risk Finance*, *3*, 57-71. <https://doi.org/10.1108/eb043483>
- Consigli, G., & Dempster, M. A. H. (1998). The CALM Stochastic Programming Model for Dynamic Asset-Liability Management. In W. T. Ziemba, & J. M. Mulvey (Eds.), *Worldwide Asset and Liability Modeling* (pp. 464-500). Cambridge University Press. <https://doi.org/10.2139/ssrn.34780>
- Dert, C. (1995). *Asset Liability Management for Pension Funds: A Multistage Chance Constrained Programming Approach*. PhD Thesis, Erasmus University.
- Dupačová, J., & Polívka, J. (2009). Asset-Liability Management for Czech Pension Funds Using Stochastic Programming. *Annals of Operations Research*, *165*, 5-28. <https://doi.org/10.1007/s10479-008-0358-6>
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*, 1-26. <https://doi.org/10.1214/aos/1176344552>
- Fleten, S. E., Høyland, K., & Wallace, S. W. (2002). The Performance of Stochastic Dynamic and Fixed Mix Portfolio Models. *European Journal of Operational Research*, *140*, 37-49. [https://doi.org/10.1016/S0377-2217\(01\)00195-3](https://doi.org/10.1016/S0377-2217(01)00195-3)
- Fourer, R., Gay, D. M., & Kernighan, B. W. (1990). A Modeling Language for Mathematical Programming. *Management Science*, *36*, 519-554. <https://doi.org/10.1287/mnsc.36.5.519>
- Høyland, K., & Wallace, S. W. (2001). Generating Scenario Trees for Multistage Decision Problems. *Management Science*, *47*, 295-307. <https://doi.org/10.1287/mnsc.47.2.295.9834>
- Kaut, M., & Wallace, S. W. (2011). Shape-Based Scenario Generation Using Copulas. *Computational Management Science*, *8*, 181-199. <https://doi.org/10.1007/s10287-009-0110-y>
- Klein Haneveld, W. K., Streutker, M. H., & van der Vlerk, M. H. (2010). An ALM Model for Pension Funds Using Integrated Chance Constraints. *Annals of Operations Research*, *177*, 47-62. <https://doi.org/10.1007/s10479-009-0594-4>
- Kouwenberg, R. (2001). Scenario Generation and Stochastic Programming Models for Asset Liability Management. *European Journal of Operational Research*, *134*, 279-292. [https://doi.org/10.1016/S0377-2217\(00\)00261-7](https://doi.org/10.1016/S0377-2217(00)00261-7)
- Mettler, U. (2005). *Projecting Pension Fund Cash Flows*. Technical Report 1, National Centre of Competence in Research Financial Valuation and Risk Management.
- Mulvey, J. M., Gould, G., & Morgan, C. (2000). An Asset and Liability Management System for Towers Perrin-Tillinghast. *Interfaces*, *30*, 96-114. <https://doi.org/10.1287/inte.30.1.96.11617>
- Nathan, H. (2016). *Detection Probability of Invasive Ship Rats: Biological Causation and Management Implications*. PhD Thesis, University of Auckland.
- Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science and Business Media.
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*, 507-554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>

- Saudi Stock Exchange (n.d.). <https://www.tadawul.com.sa>
- Schwaiger, K. (2009). *Asset and Liability Management under Uncertainty: Models for Decision Making and Evaluation*. PhD Thesis, Brunel University, School of Information Systems, Computing and Mathematics.
- Sheikh Hussin, S. A. (2012). *Employees Provident Fund (EPF) Malaysia: Generic Models for Asset and Liability Management under Uncertainty*. PhD Thesis, Brunel University, School of Information Systems, Computing and Mathematics.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48, 1-48. <https://doi.org/10.2307/1912017>
- Sklar, M. (1959). Fonctions de Repartition an dimensions et Leurs Marges. *Publications de l'Institut de Statistique de l'Universit'e de Paris*, 8, 229-231.
- Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23, 1-46. <https://doi.org/10.18637/jss.v023.i07>
- The General Organization for Social Insurance (GOSI) (n.d.). <http://www.gosi.gov.sa>
- Trading Economics (n.d.). <http://www.tradingeconomics.com/saudiarabia/interest-rate>  
<http://www.gosi.gov.sa>
- Vázsonyi, M. (2006). Overview of Scenario Tree Generation Methods Applied in Financial and Economic Decision Making. *Periodica Polytechnica. Social and Management Sciences*, 14, 29-37. <https://doi.org/10.3311/pp.so.2006-1.04>
- Whitmore, G. A., & Findlay, M. C. (1978). *Stochastic Dominance: An Approach to Decision-Making under Risk*. Lexington Books.
- Young, M. R. (1998). A Minimax Portfolio Selection Rule with Linear Programming Solution. *Management Science*, 44, 673-683. <https://doi.org/10.1287/mnsc.44.5.673>

## Appendix: The Algebraic Formulation of ALM Optimisation Models

In what follows, we present the formulations of the four ALM optimisation models described in Section 4. All models are two-stage stochastic programs with a scenario tree in the form of a fan. The following notations are used:

$I$  = The number of financial assets available for investment.

$T$  = The number of time periods.

$S$  = The number of scenarios.

**The parameters** of the model are:

$OP_i$  = The amount of money held in asset  $i$  at the initial time period  $t = 0$ ;  $i = 1, \dots, I$ .

$L_0$  = Aggregated liability payments to be made “now” ( $t = 0$ ).

$C_0$  = The funding contributions received “now” ( $t = 0$ ).

$L_{t,s}$  = Liability value for time period  $t$  under scenario  $s$ ,  $t = 1, \dots, T$ ,  $s = 1, \dots, S$ .

$C_{t,s}$  = The contributions paid into the fund at time period  $t$  under scenario  $s$ ,  $t = 1, \dots, T$ ,  $s = 1, \dots, S$ .

$R_{i,t,s}$  = The rate of return of asset  $i$  at time period  $t$  under scenario  $s$ ,  $i = 1, \dots, I$ ,  $t = 1, \dots, T$ ,  $s = 1, \dots, S$ .

$u_i$  = The upper bound imposed on the investment in asset  $i$ ,  $i = 1, \dots, I$ .

$\psi$  = The transaction cost expressed as a percentage of the value of each trade.

$\pi_s$  = The probability of scenario  $s$  occurring;  $s = 1, \dots, S$ .

$\theta$  = The maximum value of expected shortfall of the funding ratio with respect to target 1.1.

$d > 0$  = Desired rate of return over the investment horizon.

$\lambda = 1.1$  the target funding ratio.

Let us denote the **first stage decision variables** by:

$B_{i,0}$  = The monetary value of asset  $i$  to buy at the beginning of the planning horizon ( $t = 0$ );  $i = 1, \dots, I$ .

$S_{i,0}$  = The monetary value of asset  $i$  to sell at  $t = 0$ ;  $i = 1, \dots, I$ .

$H_{i,0}$  = The monetary value of asset  $i$  to hold at  $t = 0$ ;  $i = 1, \dots, I$ .

with  $H_{i,0} = OP_i + B_{i,0} - S_{i,0}$ ,  $i = 1, \dots, I$ .

**Recourse decision variables:**

$B_{i,t,s}$  = The monetary value of asset  $i$  to buy at time  $t$  under scenario  $s$ ,  $i = 1, \dots, I$ ,  $t = 1, \dots, T-1$ ,  $s = 1, \dots, S$ .

$S_{i,t,s}$  = The monetary value of asset  $i$  to sell at time  $t$  under scenario  $s$ ,  $i = 1, \dots, I$ ,  $t = 1, \dots, T-1$ ,  $s = 1, \dots, S$ .

$H_{i,t,s}$  = The monetary value of asset  $i$  to hold at time  $t$  under scenario  $s$ ,  $i = 1, \dots, I$ ,  $t = 1, \dots, T$ ,  $s = 1, \dots, S$ .

$A_{t,s}$  = The assets value at time  $t$  under scenario  $s$ , **before** portfolio rebalancing.

**The ICCP model:**

For the ICCP model, there additional variables  $Sh_{t,s} \geq 0$ ,  $t = 1, \dots, T$ ;  $s = 1, \dots$

representing the shortfalls of the funding ratio with respect to the target  $\lambda$ .

$$\text{Max} \sum_{s=1}^S \pi_s A_{T,s}$$

**Subject to:**

- **Asset Value Constraints**

$$A_{1,s} = \sum_{i=1}^I H_{i,0} R_{i,1,s}, \quad s = 1, \dots, S$$

$$A_{t,s} = \sum_{i=1}^I H_{i,t-1,s} R_{i,t,s}, \quad t = 2, \dots, T; s = 1, \dots, S$$

- **Asset Holding Constraints**

$$H_{i,0} = OP_i + B_{i,0} - S_{i,0}, \quad i = 1, \dots, I$$

$$H_{i,1,s} = H_{i,0} R_{i,1,s} + B_{i,1,s} - S_{i,1,s}, \quad i = 1, \dots, I; s = 1, \dots, S$$

$$H_{i,t,s} = H_{i,t-1,s} R_{i,t,s} + B_{i,t,s} - S_{i,t,s}, \quad i = 1, \dots, I; t = 2, \dots, T-1; s = 1, \dots, S$$

$$H_{i,T,s} = H_{i,T-1,s} R_{i,T,s}, \quad i = 1, \dots, I; s = 1, \dots, S$$

- **Fund Balance Constraints**

$$\sum_{i=1}^I B_{i,0} (1 + \psi) + L_0 = \sum_{i=1}^I S_{i,0} (1 - \psi) + C_0$$

$$\sum_{i=1}^I B_{i,t,s} (1 + \psi) + L_{t,s} = \sum_{i=1}^I S_{i,t,s} (1 - \psi) + C_{t,s}, \quad t = 1, \dots, T-1; s = 1, \dots, S$$

- **Short-Selling Constraints**

$$S_{i,0} \leq OP_i, \quad i = 1, \dots, I$$

$$S_{i,t,s} \leq H_{i,t-1,s}, \quad i = 1, \dots, I; t = 1, \dots, T-1$$

- **Bound Constraints**

$$H_{i,t,s} \leq u_i \sum_{i=1}^I H_{i,t,s}, \quad i = 1, \dots, I; t = 1, \dots, T; s = 1, \dots, S$$

- **The Integrated Chance Constraint**

$$A_{t,s} - \lambda L_{t,s} + S h_{t,s} \geq 0, \quad t = 1, \dots, T; s = 1, \dots, S$$

$$\sum_{s=1}^S \pi_s S h_{t,s} \leq \theta, \quad t = 1, \dots, T$$

$$S h_{t,s} \geq 0, \quad t = 1, \dots, T; s = 1, \dots, S$$

In our computational work, we used a variant of this model in which the objective is to minimise the expected shortfall of the funding ratio at first time, with a constraint of a minimum terminal wealth.

**The Maximin model:** For the Maximin model, there is an additional free variable  $\delta$  representing the lowest outcome of the funding ratio at stage one.

$$\text{Max} \quad \delta$$

**Subject to:**

- **Asset Value Constraints**

$$A_{1,s}/L_{1,s} \geq \delta, \quad s = 1, \dots, S$$

- **Expected terminal wealth constraint**

$$\sum_{s=1}^S \pi_s A_{T,s} \geq \sum_{i=1}^I OP_i (1+d)$$

The Asset Value Constraints, Asset Holding Constraints, Fund Balance Constraints, Short-Selling Constraints and Bound Constraints, formulated above, also hold.

**The SSD Scaled model**

For the SSD models, additional parameters specify a target distribution for the distribution of the first stage funding ratio:  $asp_k, k = 1, \dots, S$  are aspiration levels for the means of the worst  $k$  outcomes of the funding ratio distribution,  $k = 1, \dots, S$ .

This

**The additional variables for the SSD models are:**

$F_s$  = The funding ratio under scenario  $s$  at time  $t = 1$ ; ( $F_s = A_{1,s}/L_{1,s}$ );  $s = 1, \dots, S$

$T_k$  = The  $k$ -th worst outcome of the funding ratio at time 1,  $k = 1, \dots, S$  (free variable); thus,  $T_1, \dots, T_S$  are the outcomes of a random variable equal in distribution to the funding ratio

$Z_k$  = The mean of the worst  $k$  outcomes of the funding ratio, or other said, ScaledTail $_{k/S}(F)$ ;  $Z_k = (T_1 + \dots + T_k)/k$ ,  $k = 1, \dots, S$  (free variable)

$\delta = \min_{k=1, \dots, S} (Z_k - asp_k)$  = the worst partial achievement (free variable);

$d_{k,s}$  = Non-negative variables,  $d_{k,s} = [T_k - F_s]^+$  that is

$$d_{k,s} = \begin{cases} 0, & \text{if } F_s \geq T_k \\ T_k - F_s, & \text{otherwise} \end{cases}$$

The objective is to maximise the minimum difference between the *mean* of the worst  $k$  funding ratios at time  $t = 1$  and the  $k$ -th aspiration level; a regularisation term is added to tackle the case of multiple optimal solutions, with  $\epsilon$  a positive number close to 0.

$$\text{Max } \delta + \epsilon \left( \sum_{k=1}^S Z_k - \sum_{k=1}^S asp_k \right)$$

**Subject to**

- **Funding Ratio Definition**

$$F_s = \sum_{i=1}^I H_{i,0} R_{i,1,s} / L_{1,s} \quad (F_s = A_{1,s} / L_{1,s}), \quad s = 1, \dots, S$$

- **Additional Constraints to Formulate the SSD Model**

$$Z_k = T_k - \frac{1}{k} \sum_{s=1}^S d_{k,s}, \quad k = 1, \dots, S$$

$$Z_k - asp_k \geq \delta, \quad k = 1, \dots, S$$

$$T_k - F_s \leq d_{k,s}, \quad k, s = 1, \dots, S$$

$$d_{k,s} \geq 0, \quad k, s = 1, \dots, S$$

- **Terminal Wealth Constraint**

$$\frac{1}{S} \sum_{s=1}^S A_{T,s} \geq \sum_{i=1}^I OP_i (1+d)$$

The Asset Value Constraints, Asset Holding Constraints, Fund Balance Constraints, Short-Selling Constraints, and Bound Constraints, formulated above, also hold.

The SSD unscaled model is obtained in a similar way, by considering aspiration levels for *sums* (instead of means) of the worst outcomes of the funding ratio. For the complete formulation and more details, please see [Alwohaibi & Roman \(2018\)](#).

In all models, the right-hand side of the constraint on the expected terminal asset value is the same (equal to  $A_T$  which corresponds to a cumulated terminal wealth of 581.5548 billion Saudi Riyals (SAR)). The value of  $\varepsilon$  is fixed at 0.0001.