Report I - SF2980 Risk Management

Anna Gerbens 991015-T381 Luis Hausladen 001111-T275
Samy Harafa 991002-T352 Baptiste Le Roux 990909-T556

22 November, 2022

Project Title

Objectives

In this project we evaluate the value of pension savings, including its risk measures, over a 30-year period by considering a yearly investment of 1,000 dollars in long positions in a portfolio of stocks and a risk-free one-year zero coupon bond. We do this for three different scenarios.

First of all, we consider the scenario where Sweden's pension system is based on fund insurance and so the risky portfolio of stocks can be viewed as a portfolio selected by the customer. The yearly returns on the portfolio of stocks year k is modeled as $R_k = e^{\mu+\sigma Z_k}$ where Z_k is standard Normally distributed. The yearly returns are assumed to be independent. The yearly return on the risk-free bond is assumed to be $e^{0.01}$. For each year k, after adding the investment of 1,000 dollars, the entire portfolio is rebalanced such that the value invested in the stock at the beginning of year k is a specified fraction g_k that depends on the year k and the parameters p,$c \in [0,1]$.

The first objective is to to determine a function f that expresses the value of the pension savings in 30 years by using the function parameters p, c, the random variable Z_k, μ and σ. Then, we determine and compare the empirical distribution of the value of the pension in 30 years for varying values of p, c, μ and σ. To do so, we simulate a sample of suitable size n from the distribution of Z_k and use a recursive formula for computing the portfolio value.

The second objective is to suggest a suitable criterion of optimality for a portfolio and determining empirical optimal values of p and c for fixed μ and σ. To do so, we perform a gridsearch. The mean and the 1% quantile of this 'optimal portfolio' also need to be reported.

In the second scenario, Sweden's pension system is based on a traditional life insurance system where all investments have a guaranteed annual return that is slightly lower than the risk-free rate. The fund manager of the life insurance policy then invests the necessary amount in the risk-free bond to cover the guarantee and the remaining surplus in the risky portfolio of stocks to generate a higher return. Here, the objective is to compare the values of the pension in 30 years using this strategy for a fixed guaranteed annual return with the values of the pension in 30 years using the fund insurance strategy. The mean and the 1% quantile of the portfolio in the traditional life insurance scenario need to be reported.

In the third scenario, using the traditional life insurance system, a leverage is added in the exposure to stocks which will generate larger profits. This type of strategy is often referred to as a constant proportion portfolio insurance (CPPI). The objective is to compare the values of the pension in 30 years using this strategy with the values obtained using the fund insurance and the traditional life insurance scenarios. Again the mean and the 1% quantile of this portfolio need to be reported.

Finally, instead of using the log-normal distribution for the annual returns of the portfolio of stocks in the three scenarios, we repeat the analysis using an empirical distribution based on monthly historical data.

Mathematical Background

0.0.1 Returns

Let V_k denote the value of a portfolio at the beginning of year k and R_k the return of that portfolio during year k, then the value of portfolio at the start of the following year is:

$$V_{k+1} = V_k \ R_k \tag{1}$$

For the assignments a) to d), we assume that R_k is of the form

$$R_k = e^{\mu + \sigma Z_k}$$

where Z_k has standard normal distribution.

0.0.2 Order Statistics

Given n samples $(L_1, ..., L_n)$ from a random variable L, the corresponding order statistics are $L_{1,n} \geq L_{2,n} \geq ... \geq L_{n,n}$ i.e the ordered samples.

0.0.3 Cumulative Distribution Function - CDF

Let X be a random variable, its cumulative distribution function (often shortened as CDF) is:

$$F_X(x) = P(X \leq x) \tag{2}$$

Domains of smallest and greatest values of x are called the left and right tails of the distribution

The quantile at probability level p of X is:

$$F_X^{-1}(p) = inf\{\ x \mid P(X \leq x) \geq p\} \tag{3}$$

By sampling n independant and identically distributed samples $x_1, ..x_n$ from X we can compute the empirical CDF of X which we denote $F_n(x)$ by assigning weight of $\frac{1}{n}$ to each sample resulting in a piecewise constant function with values between 0 and 1:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x \geq x_i) \tag{4}$$

0.0.4 Risk measures

Throughout this project two risk measures will be used, value-at-risk (at probability level p, VaR_p) and expected shortfall (at probability level p, ES_p). Let X be a random variable, its VaR and ES are defined as follows:

$$VaR_p(X) = inf\{m : P(X + mR_0 < 0) \leq\ p\ \} \tag{5}$$

VaR can be interpreted as "One will not loose more than VaR with probability p"

$$ES_p(X) = \frac{1}{p} \int_0^p VaR_u(X)du \tag{6}$$

Unlike VaR, expected shortfall captures how much one will loose should they loose more than VaR_p. That is, ES captures the amount of risk existing in the tail of the distribution. ES_p can be interpreted as "How much one will loose knowing that they loose more than VaR_p"

0.0.5 Loss function

Let X a random variable denoting the net worth of a portfolio. Given some risk-free rate R_0 the loss function L is defined as:

$$L = -X/R_0 \tag{7}$$

and related to VaR as follows:

$$\begin{aligned} VaR_p(X) &= inf\{m : P(X + mR_0 < 0) \leq \ p \ \} \\ &= inf\{m : P(X < -mR_0) \leq \ p \ \} \\ &= inf\{m : P(-X/R_0 > m) \leq \ p \ \} \\ &= inf\{m : 1 - P(-X/R_0 \leq m) \leq \ p \ \} \\ &= inf\{m : P(L \leq m) \geq \ 1 - p \ \} \\ &= F_L^{-1}(1 - p) \end{aligned} \tag{8}$$

In the case of empirical cdf with n samples, the loss function can be related to VaR and ES as follows:

$$VaR_p(X) = L_{[np]+1,n} \tag{9}$$

$$ES_p(X) = \frac{1}{p} \sum_{k=0}^{[np]} \frac{L_{k,n}}{n} + L_{[np]+1,n}(p - \frac{[np]}{n}) \tag{10}$$

Results

a) Proof by induction of V_{30} formula

Let V_k the overall value at the start of year k.
Initially, we have $V_0 = 0$

Let $g_k = p(1 - c(k-1)/30)$ the fraction of the overall value invested in the stock portfolio for year k.
Then V_k is related to V_{k-1} as follows:

$$V_k = (V_{k-1} + 1000)(g_k \ R_k + (1 - g_k) \ e^{0.01}) = (V_{k-1} + 1000) \ A_k \tag{11}$$

where

$$A_k = (g_k \ R_k + (1 - g_k) \ e^{0.01}) \tag{12}$$

We will proceed to proove induction hypothesis H_k: V_k can be explicitly written (in the general case where $V_0 \neq 0$)

$$V_k = V_0 \ (\prod_{i=1}^{k} A_j) + 1000 \ \sum_{j=1}^{k} (\prod_{i=j}^{k} A_i) \tag{13}$$

Base case:
V_1 is obtained by adding 1000$ to initial capital V_0 splitting between stock and ZBC as g_1 and $(1 - g_1)$ and earning the returns of one year of interests. That is:

$$\begin{aligned} V_1 &= (V_0 + 1000)(g_1\ R_1 + (1 - g_1)\ e^{0.01}) \\ &= V_0\ A_1 + 1000\ A_1 \end{aligned} \tag{14}$$

Thus H_1 holds.

Induction:
Suppose H_k holds with $k > 1$, let's proove that $H_{k+1} does$

$$\begin{aligned} V_{k+1} &= (V_k + 1000)\ A_{k+1} \\ &= ((V_0\ (\prod_{i=1}^{k} A_j) + 1000\ \sum_{j=1}^{k}(\prod_{i=j}^{k} A_i)) + 1000)\ A_{k+1} \\ &= ((V_0\ (\prod_{i=1}^{k+1} A_j) + 1000\ \sum_{j=1}^{k}(\prod_{i=j}^{k+1} A_i)) + 1000\ A_{k+1}) \\ &= V_0\ (\prod_{j=1}^{k+1} A_j) + 1000\ \sum_{j=1}^{k+1}(\prod_{i=j}^{k+1} A_i) \end{aligned} \tag{15}$$

Thus H_{k+1} holds.

Conclusion:
For any positive integer k, H_k holds. In particular for $V_0 = 0$ and $k = 30$:

$$\begin{aligned} V_{30} &= 1000\ \sum_{j=1}^{30}(\prod_{i=j}^{30}(g_i\ R_i + (1 - g_i)\ e^{0.01})) \\ V_{30} &= 1000\ \sum_{j=1}^{30}(\prod_{i=j}^{30}(p(1 - c(i-1)/30)\ e^{\mu + \sigma Z_k} + (1 - p(1 - c(i-1)/30))\ e^{0.01})) \end{aligned} \tag{16}$$

0.0.6 a) Empirical cdf of V_{30} for different parameter values

Based on a sample $\{Z_1, \ldots, Z_n\}$ of size $n = 1000$, we computed the empirical cdf of V_{30} with different parameter values. We changed only one parameter at once to investigate the influence of each parameter seperately. While one parameter is varied, the others stay constant at the default values

$$\mu = 0.03, \qquad \sigma = 0.2, \qquad p = 0.5, \qquad c = 0.5.$$

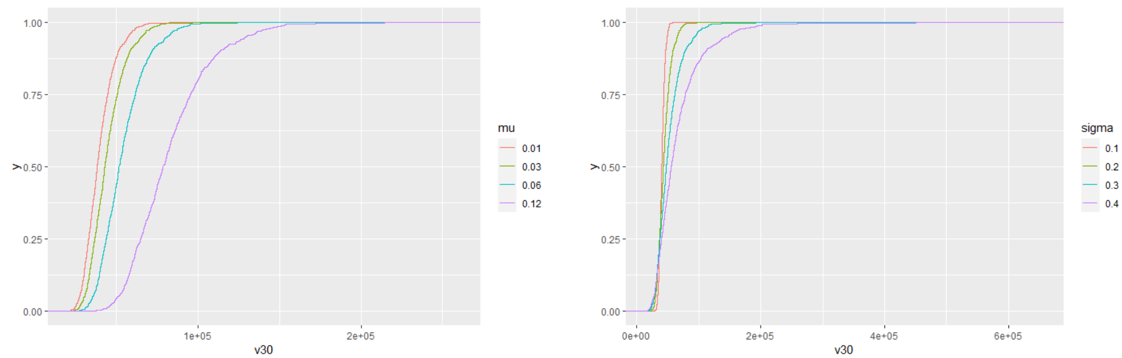

Figure 1: Empirical distribution function of V_{30} for different values of μ and σ based on a sample of size $n = 1000$.

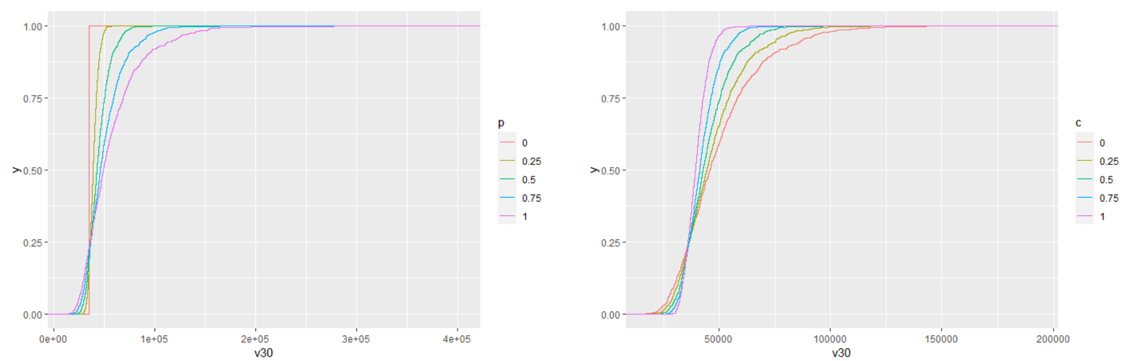

Figure 2: Empirical distribution function of V_{30} for different values of p and c based on a sample of size $n = 1000$.

Figure 1 : Mean and standard deviation of R_k

From **Figure 1** we can see that unsurprisingly, increasing the expected value of the log normal distribution of R_k shifts the empirical cdf of V_{30} to the right, that is, the probability of lower values of V_{30} decreases as μ increases. Still from **Figure 1** we observe that consistently with our expectations, increasing standard deviation σ slightly increases the probability of lower values of V_30 but also increases the probability of higher values of V_30 indeed the purple curve with higher sigma is above other curves before a crossing point around 40 000, but under all other curvers after that point meaning that corresponding V_{30} samples are distributed more towards high values.

Figure 2 : Balance of the portfolio

We recall that the every year the portfolio is rebalanced as $g_k = p(1 - c(k-1)/30)$ in the risky asset. From that formula we see that p is the initial proportion dedicated to the risky asset and c the change rate of portfolio balance over the years. From **Figure 2** we can see that higher values of p increases the risk as a counterweight for higher possible returns, when p=0 portfolio is fully deterministic and the cdf looks like a step function at $V 30$. Higher values of c corresponding to faster rebalance to the riskless asset over the years reduces risk (we can see empirical cdf compressed towards its center) while for c=0 which corresponds to constant share of risky asset in portfolio expands the cdf towards extremal values allowing prospective higher returns but also higher losses, that is, more risk.

0.0.7 b) Criterion for optimality

In this section, we calculate the empirical expected shortfall with formula (7.11) on page 212 in the coursebook. For this, we need the loss, which we calculate as follows: Let $X = W_1 - W_0 R$, where $W_1 = V_{30}$ is the final value of the portfolio and

$$W_0 R = 1000 \cdot \sum_{k=1}^{30} e^{kr},$$

is the guaranteed value, we could have achieved with the same amount of money by investing 1000 euros in the risk free asset each year. Now X models the net worth of our portfolio at year 30 and the loss is defined as:

$$L = -\frac{X}{R} = -X \cdot e^{-30r},$$

i.e., the value of $-X$ discounted to today.

For the effects of changing p and c on the ecdf, when $\mu = 0.03$ and $\sigma = 0.2$, see the previous subsection. In order to measure quantitatively the effects of changing p and c on the ecdf, we computed empirical estimates of the expected value $E[V_{30}]$, the expected shortfall $ES_{0.01}$ and the 1% quantile $F_{V30}^{-1}(0.01)$ for different values of p and c, using a sample of size $n = 1000$. If we plot these three indicators for different values of p and c, we obtain Figure 3.

Note that the larger p and the smaller c, the riskier our portfolio since we invest a higher percentage in the stock. With this in mind, the following observations from Figure 3 make complete sense: The quantile $F_{V30}^{-1}(0.01)$ decreases when p increases (with c fixed) and increases when c increases (with p fixed), whereas the mean $E[V_{30}]$ and expected shortfall $ES_{0.01}$ behave opposite: They are smallest for $p = 0$ (then the coice of c is irrelevant) and largest for $p = 1, c = 0$ (corresponding to the most risky portfolio). Since we want $E[V_{30}]$ as high as possible, but $ES_{0.01}$ as low as possible, a trade off approach is required to find an optimal portfolio.

Our idea for the trade off approach is to evaluate the ratio between $E[V_{30}]$ and $ES_{0.01}$. Since the fraction $\frac{E[V_{30}]}{ES_{0.01}}$ is high when $E[V_{30}]$ is high and $ES_{0.01}$ is low, we search for pairs

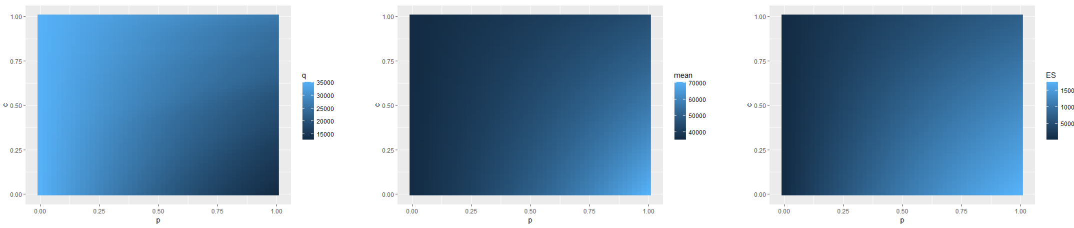

Figure 3: $F_{V_{30}}^{-1}(0.01)$, $E[V_{30}]$, and $ES_{0.01}$ for $p, c \in \{0, 0.02, 0.04, \ldots, 0.98, 1\}$, based on a sample size of $n = 1000$.

(p, c) maximising this fraction. However, since $ES_{0.01} \to 0$ for $p \to 0$ and $E[V_{30}] \to C > 0$ for $p \to 0$ ($E[V_{30}]$ converges to the value we get when putting everything in the bank account, i.e. p=0), the fraction will get arbitrarily large for small values of p (see left plot im Figure 4). So we need another restriction to make a reasonable coice of (p, c). We decided on the following: The pair (p,c) should corresond to a portfolio whose mean is at least 60% of the maximum possible mean (the mean for $p = 1$ and $c = 0$). Like this, we ensure a rather high mean and we are able to find reasonable pairs (p,c) maximising the fraction. The right plot in Figure 4 displays the values of the fraction after restricting to pairs yielding at least 60% of the highest mean.

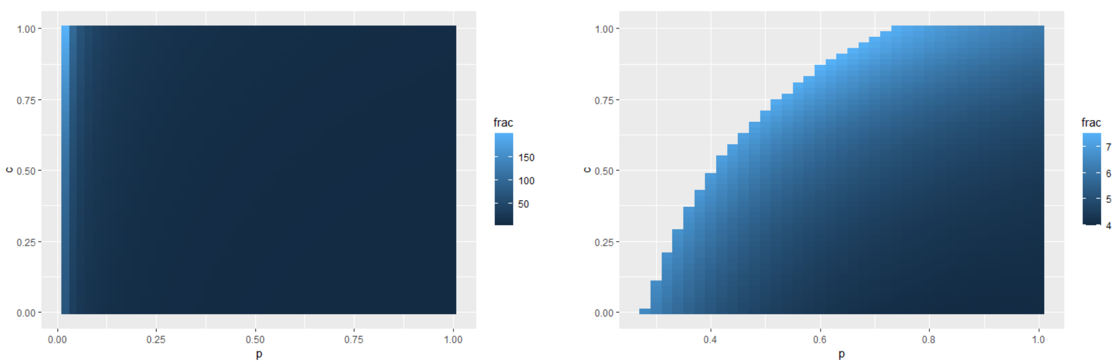

Figure 4: Value of the fraction $\frac{E[V_{30}]}{ES_{0.01}}$ for $p, c \in \{0, 0.02, 0.04, \ldots, 0.98, 1\}$, $p > 0$, based on a sample size of $n = 1000$.

0.0.8 b) Optimal values of (p, c)

The right plot in Figure 4 provides all information we need to retrieve the optimal value (p, c). But since the values on the borderline are quite similar, it is hard to read off the exact pair just by visual inspection. After inspecting the raw data, we see that the tuple

$(p, c) = (0.60, 0.86)$ yields the highest fraction. Using these values, we get

$$E[V_{30}] = 42380, \qquad F_{V_{30}}^{-1}(0.01) = 28187.$$

However, we realised that these values are sensible to the random initialisation of the Z_k (i.e. the "situation on the market in the future"). To get a more stable estimate of the *real* optimum (p, c), we increased our sample size to $n = 10,000$ and computed the optimal (p, c) in 10 simulations. Our estimate for the true optimum is now the mean of these 10 simulations:

$$(\widehat{p^*}, \widehat{c^*}) = (0.683, 0.917)$$

If we use $(\widehat{p^*}, \widehat{c^*})$ for the data used in Figure 4, we get:

$$E[V_{30}] = 42812, \qquad F_{V_{30}}^{-1}(0.01) = 27802.$$

0.0.9 c) Recursive formula for V_k in traditional life insurance

With the notation and description from the assignment, we get the following recursive formula for V_k:

$$\begin{aligned} V_k &= (V_{k-1} + 1000 - G_k e^{-r}) \cdot R_k + G_k e^{-r} \cdot e^r \\ &= (V_{k-1} + 1000 - G_k e^{-r}) \cdot R_k + G_k, \end{aligned} \tag{17}$$

and $V_0 = 0$.

0.0.10 c) Histogram of simulated values and comparison to results in b)

Implementing (17) in R and simulating $n = 1000$ samples of V_{30}, yields the histogram in Figure 5. For the mean and 1% quantile of V_{30}, we get the empirical estimates:

$$E[V_{30}] = 36835, \qquad F_{V_{30}}^{-1}(0.01) = 33551$$

The expected value is a lot lower than in b), but also the 1%-quantile is a lot higher than in b). This means that traditional life ensurance is less risky but also produces less returns on average.

0.0.11 d) Recursive formula for V_k in traditional life insurance with leverage

Let α denote the leverage. With the notation and description from the assignment, we get the following recursive formula for V_k:

$$\begin{aligned} V_k &= \alpha(V_{k-1} + 1000 - G_k e^{-r}) \cdot R_k + (V_{k-1} + 1000 - \alpha(V_{k-1} + 1000 - G_k e^{-r}) \cdot e^r \\ &= \alpha(V_{k-1} + 1000 - G_k e^{-r}) \cdot R_k + (1 - \alpha)(V_{k-1} + 1000) e^r + \alpha G_k, \end{aligned}$$

if $V_k \geq G_k$ ($\alpha = \frac{1}{0.5} = 2$ for this assignent) and

$$V_k = (V_{k-1} + 1000) \cdot e^r,$$

if $V_k < G_k$ (i.e. the portfolio falls behind the guarantee). As above, $V_0 = 0$.

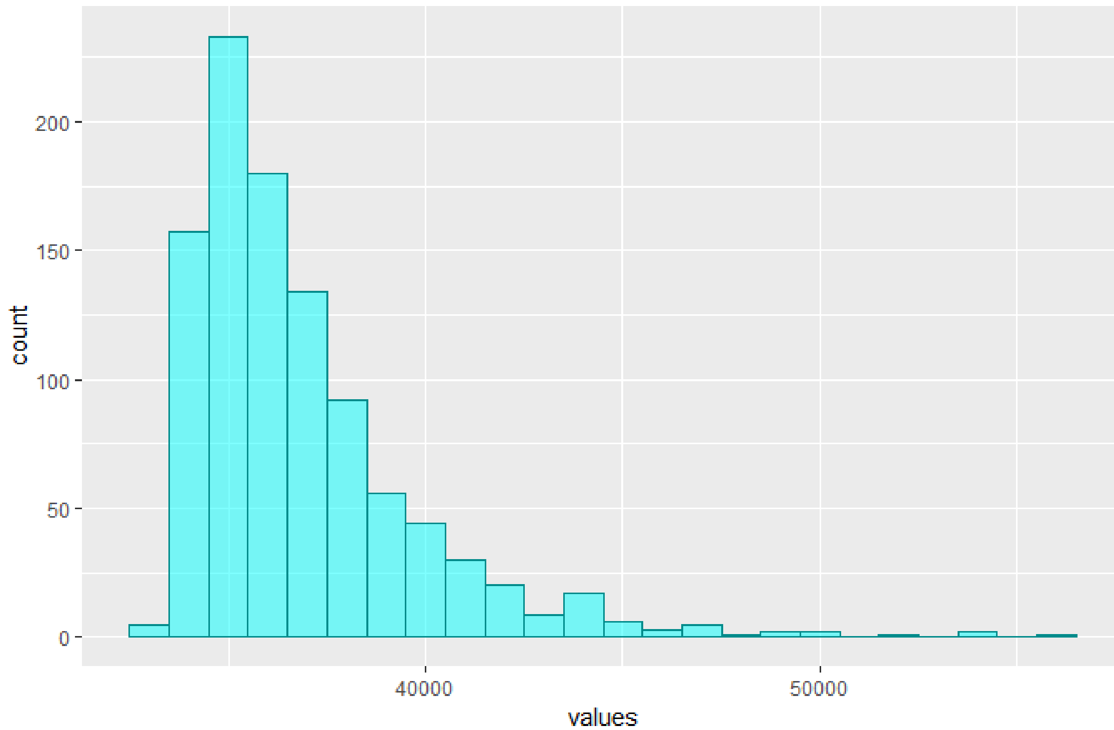

Figure 5: Histogram of V_{30} using traditional life ensurance with no leverage, based on a sample of size $n = 1000$.

0.0.12 d) Histogram of simulated values and comparison to results in b) and c)

Implementing the above formula in R and simulating $n = 1000$ samples of V_{30}, yields the histogram in Figure 6. For the mean and 1% quantile of V_{30}, we get the empirical estimates:

$$E[V_{30}] = 39767, \qquad F_{V_{30}}^{-1}(0.01) = 32815$$

We notice that the expected value is higher than with no leverage (see c)), but the 1% quantile is smaller. This is also apparent in Figure 6, where we see that using leverage yields a few very high performing portfolios (heavy right tail), but also has more observatios at the lower end of portfolio values. The expected value is higher than in c) but still a fair bit lower than in b). So if one wants to achieve the highest mean return and is not concerened about risk, the trading strategy in b) is preferred over traditional life insurance.

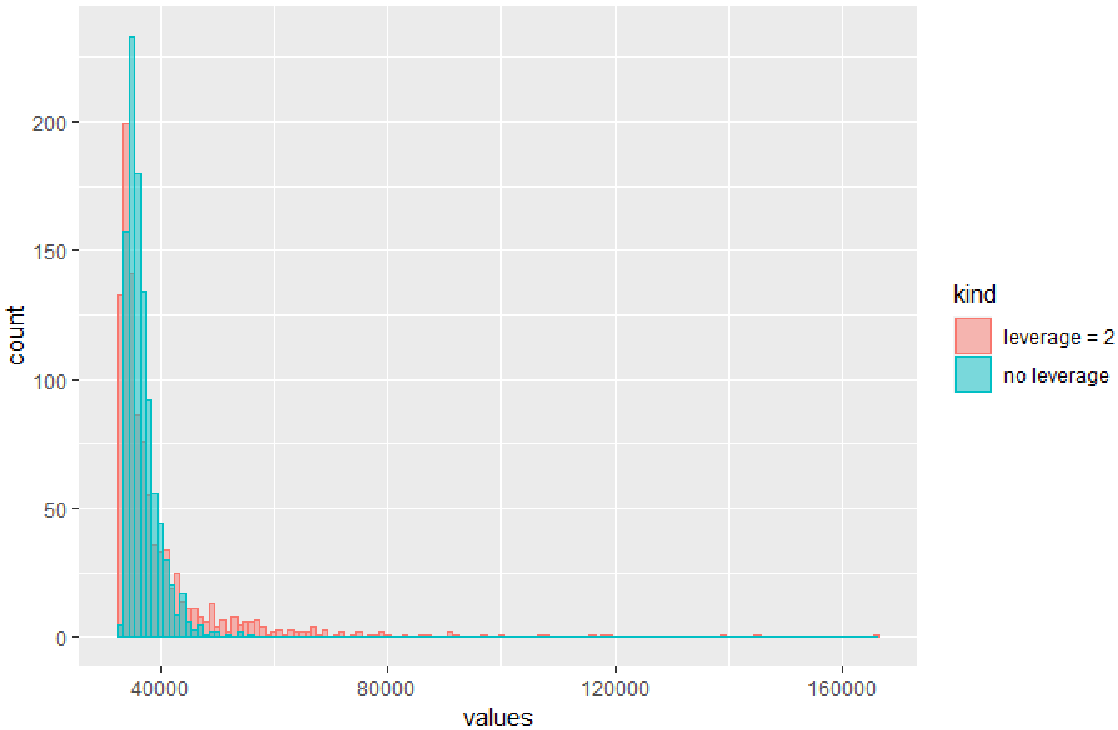

Figure 6: Comparison of the empirical distribution of V_{30} with no leverage againt a leverage of 2, based on a sample of size $n = 1000$.

0.0.13 e) Acquiring market data

From the website `https://www.pensionsmyndigheten.se/nyheter-och-press/nyheter-fondtorg/statistik-om-premiepensionens-fonder`, we downloaded the file *Statistik premiepension 2022-10-31* which can be found in the subsection *Månadsstatistik fonder och fondsparande 2022*. Of the different worksheets contained in this Excel file, we used the data on fond movement provided in the worksheet named *Index, fondrörelsen.* It contains the daily fond value from the 13th of December 2000 to the 31st of October 2022. After loading this data into `R`, we restricted to observations from the first of each month, to get monthly values of the fond and transformed them into historical monthly return rates.

e) Method of historical simulation To simulate a yearly return rate of the fond, we drew 12 monthly return rates uniformly at random with replacement from the set of historical return rates and multiplied them together to get a simulated yearly return. In the coursebook, this is referred to as *historical simulation* and is explained in *Example 7.3*. However, this approach must be used with care, because we use data from 21 years to generate returns for 30 years.

0.0.14 e) Repetition of b) to d) with market data

After generating 30 samples of yearly log returns $\{h_1, \dots, h_{30}\}$ with the procedure outlined above, we can generate one observation of V_{30} using historical market data and the same recursive formulas vor V_k as above, while replacing

$$R_k = e^{\mu + \sigma z_k}$$

with

$$R_k = e^{h_k}.$$

Repeating this n times, we get a whole sample of of V_{30} and can repeat our analysis in b), c) and d) (we leave the risk free and guaranteed rate unchanged at $r = 0.01$ and $\bar{r} = 0.005$ and generate a sample of size $n = 1000$).

In b) we now got the optimal values $(p, c) = (1, 0.78)$, resulting in

$$E[V_{30}] = 49956, \qquad F_{V_{30}}^{-1}(0.01) = 22314$$

if we use the estimate $(\widehat{p^*}, \widehat{c^*}) = (0.683, 0.917)$ from above, we get

$$E[V_{30}] = 42800, \qquad F_{V_{30}}^{-1}(0.01) = 27700$$

In Figure 7, we compare the empirical distribution of V_{30} for traditional life insurance with (red) and without (cyan) leverage using a histogram (corresponding to assignments c) and d)). It is noticable that the with leverage, the observations tend to lie further to the right and the distribution has a heavier tail since very high outlier portfolio values occur, which does not happen without leverage. This "right shift" of the distribution is also evident, when comparing the means in Table 1. But using no leverage yields a slightly higher 1% quantile, which means that this approach is more conservative. These observations are conform with the results of our simulation (see c) and d)).

	$E[V_{30}]$	$F_{V_{30}}^{-1}(0.01)$
b) (using $p = 1$, $c = 0.78$)	49956	22314
b) (using $p = 0.683$, $c = 0.917$)	42800	27700
c)	37986	34260
d)	45258	33405

Table 1: Results of the approaches b), c) and d) using market data

In Table 1 we see that we get the highest expectation using the trading strategy in b) combined with the pair (p,c) that is fitted to the specific simulated sample of future returns. As mentioned above, the traditional live ensurance with leverage yields a higher average return than without. From the table it is also apparent that a higher expectation

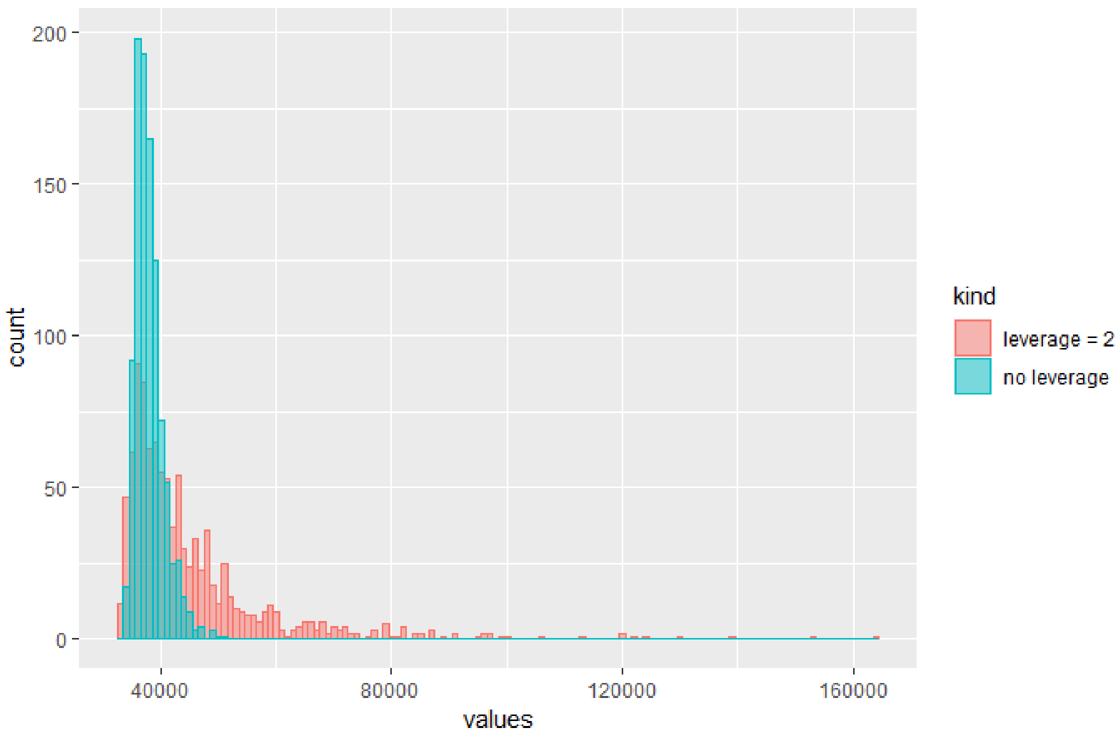

Figure 7: Comparison of the empirical distribution of V_{30} with no leverage againt a leverage of 2, based on a sample of size $n = 1000$.

means a lower 1% quantile, so there is a trade-off between both. The only thing that differs compared to not using historical data is that traditional life ensurance with leverage now outperforms b) (i.e. higher expected value *and* quantile) if we use the optimal values obtained through randomly sampling the returns. This indicates that our approach of determining an optimal portfolio is highly sensitive to the underlying market situation. In b) we fixed the average log-return of the stock to $\mu = 0.03$ and the volatility to $\sigma = 0.2$. For our generated sample of historical log-returns, we get an average log-return of 0.056 and a volatility of 0.138. So the more robust optimal values originally obtained in b) are not valid anymore and would have to be re-evalued each time, the market changes. For simplicity, we don't do that here.

Summary

In summary, one can achieve the highest mean portfolio value with the trading strategy from b) and traditional life insurance produces higher average returns with a leverage of 2 compared to no leverage. But one should be aware that there is a trade-off omnipresent: A higher expected portfolio value always yields a lower 1%-quantile. Since the portfolio is worth more than the 1%-quantile with a chance of 99%, this can be interpreted as follows: Higher expected portfolio values imply that the money we get with a probability of 99% is less. Or short: Higher rewards correspond to higher risks. The same observations were made with sampled marketdata which confirms that the model assumtion of log-normally distributed returns is a resonable model of reality.

Concluding, we recommend to use the investment strategy of traditional life ensurance because it provides a balance between high mean and high 1%-quantile.