MARY HARDY: Investment Guarantees: Modelling and risk management for equity-linked life insurance. John Wiley & Sons. ISBN 0-471-39290-1, 2003.

It is always a pleasure to read something written by Professor Hardy. It is doubly so when it is a book on a subject that I have been long concerned with, and triply so when I and two colleagues, Dr Sheauwen Yang and Professor Howard Waters, have recently presented a long paper (2003) to the actuarial bodies in the United Kingdom on a rather similar subject, that of contracts with guaranteed annuity options (GAOs). (I shall refer to this paper and its authors as "WWY"). So I am very pleased to write this review, and to be able to recommend Professor Hardy's book most warmly.

Although I might have wished that I had written a book on this subject myself, this is not exactly the book that I would or could have written. Professor Hardy's approach is very similar to that of myself and my colleagues in relation to GAOs, but it also differs in a number of places, and she describes some things she has done that I have never attempted. I shall draw attention to our differences as we go along.

The book, according to her Introduction "is designed for all practitioners working in equity-linked insurance ... It is written with actuaries in mind, but it should also be interesting to other investment professionals. [It] forms the basis for a one-semester graduate course for students of actuarial science, insurance and finance." In my view it succeeds well in these objectives. The actuarial material relating to mortality tables is tidied away into short Appendices. I am not sure that a practitioner who had no statistical or mathematical training at all could easily follow it, but it should present no difficulties to any actuary.

Equity-linked life assurance (as it is called in the U.S.A.) goes under several names: unit-linked in the United Kingdom, segregated funds in Canada. Many of the policies provide, or used to provide, guarantees of a minimum sum assured on maturity, and perhaps also on earlier death. The existence of these policies in the U.K. in the 1970s led to the seminal work done by the Maturity Guarantees Working Party (MGWP), whose report published in 1980, suggested setting up contingency reserves calculated as what are now called quantile reserves or value-at-risk reserves.

Equity-linked assurances with investment guarantees are the archetypal examples of a life insurance policy that contains benefits defined as the better of A and B, where A and B are amounts that are both defined in the policy. One can relate these to standard financial options by defining a new type of option, a Maxi, whose payoff at the expiry date is Max(A, B). A Maxi is easily related to the more usual Call and Put options. An equity-linked policy can be treated as an investment in ordinary shares ("equities" or "common stock") plus a put option, or as an investment in cash plus a call option, or as a maxi.

However, Professor Hardy's initial approach, like that of the MGWP and of WWY, is to ignore the financial option concepts, and to estimate quantile reserves (or better "conditional tail expectation" or CTE reserves) by the use of simulation. She describes this as "the actuarial approach", as opposed to the "dynamic hedging approach" of financial economists. However, many actuaries understand financial options, and many financial economists understand

the necessity for contingency reserves, so the names are no more than convenient labels.

Chapter 1 of the book describes the types of policy considered, and the history and background. To do simulations one needs a stochastic simulation model, to replicate the "real world" and this is considered in Chapter 2. In this chapter the author describes several possible models and modifications thereof. Each of the models is fitted to two data sets, monthly values from about 1956 to 2000 of the TSE 300 index and the of the S&P 500 index. The models described include the independent lognormal, autoregressive AR(1) lognormal, and regime-switching lognormal (RSLN) models for the structure, ARCH and GARCH models for the residuals, also the empirical distribution, the Wilkie model, and vector autoregressive (VAR) models. It is clear that the author prefers the RSLN model.

One must emphasise that at this point we are seeking a model to represent the real world movement of economic variables, in this case the total returns on shares. We are not concerned with option pricing models. We would like a model of the real world that is as realistic as we can make it, and we can justify from the data, without its becoming intractable for simulation. If we wish to restrict ourselves to total returns on shares then the RSLN seems to have many advantages. But in general it seems to me to be a pity to look only at total returns. Share dividends, and share earnings are additional information, to which participants in the market do pay attention. The rate of inflation and interest rates, and for a country like Canada exchange rates and what is happening in the United States, may also be relevant, as Hardy later observes (on page 87). Therefore I would prefer to use an integrated model, on the lines of the Wilkie model, rather than model restricted to one series. However, I see no reason why we should put ourselves in the straightjacket of a VAR model. The relationships between variables may not be all strictly linear.

Harris (1999) has applied RSLN models to multivariate data. Whitten & Thomas (1999) apply a threshold model to multivariate data. In both cases there are multiple regimes (but restricted in their examples to two). All variables are in the same regime at once (but one could imagine a model where this did not apply). In the RSLN model the regime switches at random between models with specified probabilities. In the threshold model the regime is in one model or another depending on the value of an indicator variable in the previous period (whether or not inflation was higher or lower than 10%). An elaboration of Hardy's RSLN model would be to include the US index and the Canadian index in one model, and define four states where neither, one or other, or both are in the higher variance regime. This could take account of the connection between the states that is observed on page 87.

To use a model we must estimate parameters for it, and the next three chapters discuss this. In Chapter 3 Professor Hardy discusses the classical maximum likelihood estimation (MLE) method, how one derives the MLE parameter values, uses the information matrix to derive confidence intervals for and correlations between the parameter estimates, and then uses criteria such as the likelihood ratio test and Akaike criterion to choose between models. This is standard material, clearly presented.

In Chapter 4 the "left tail calibration method" is described. For the particular application low share returns are critical, so it is desirable that the left (negative) tail of the distribution is adequately modelled. It is clear that the monthly returns are "fat-tailed", which is why an ARCH or GARCH or RSLN model is much better than a simple lognormal model for representing the whole distribution. But it is possible to adjust the parameters of any of the models (usually just the standard deviation) so that the left tail is adequately "fattened". This usually means that the right tail is not fitted so well. The motivation for this process is also to meet the requirements of the Canadian Institute of Actuaries' report on segregated funds, which would allow a life office to use any model it wished provided that certain statistics in relation to the left tail are adhered to.

One method that Professor Hardy does not discuss is to use a fat-tailed distribution (other than a stable distribution) for the residuals. If Z represents the standardised (0,1) residuals, one can generate Z as $X_1 - X_2$ where X_1 and X_2 are distributed with any distributions defined on $(0, \infty)$, such as lognormal, gamma, Weibull, Pareto or many others. Since X_1 dominates the right tail and X_2 the left tail, one can fit the tails separately if one wishes. The MLE method would be difficult, but one can match higher moments or quantiles. The method is mentioned by WWY and seems worth considering.

In Chapter 5 we move on to Bayesian Markov Chain Monte Carlo (MCMC) methods. I have previously found these difficult to follow, I suppose because I have not in fact implemented them myself (I do not feel that I really understand a numerical mathematical method unless I have written a computer programme to implement it), but I find Hardy's explanation as clear as any I have seen so far. The advantage of the MCMC method, which indeed looks complicated as compared with the MLE method, is that it gives empirical, simulated, distributions for the parameters. The MLE method gives the covariance matrix of the parameter estimates, but one then assumes normality, and the results are only asymptotically normal. The MCMC method shows that the distribution of the parameter estimates is not as normal as one might have hoped.

This is important when we come later to discuss the effect of parameter uncertainty on the simulation results for the investment guarantees. One can allow for this by using a different set of parameters for each simulation, picked from a multivariate distribution of the parameters, using what WWY call a "hypermodel". MLE gives a multivariate normal distribution from which one can pick. MCMC gives an empirical multivariate distribution, with as many values to pick from as one has chosen to simulate in the MCMC procedure. There are both theoretical and practical considerations that might influence which method one chooses to use. Normal distributions can be awkward if the parameters are essentially positive (such as a variance) or restricted to a range such as (0,1) or (-1,1) (such as an autoregressive parameter), but one can transform the parameter (assume that log variance is normal), or just restrict it to the desired range (set any value greater than 1 to 1).

Using an empirical distribution requires large computer storage, which may or may not be a problem, and restricts the drawn parameter values to the range in the empirical distribution. If one simulates enough values by MCMC that may not be a problem; but it may put up the storage requirements. It seems that there would be more work to be done before one could say that MCMC methods should always be used, but I am sure that they should be tried out.

In Chapter 6 Professor Hardy shows how to model the guarantee liability using "the actuarial method", that is by setting up a contingency reserve at the start of the contract, which is invested in a specific, but unchanging, way, and which has a given chance (e.g. 99%) of being sufficient to meet the emerging liability.

The contracts that she describes in Canada have some features that may not be customary elsewhere, and this complicates things. Thus the policies usually have a guaranteed minimum benefit on death and also on maturity, though these may be defined differently; but also there may be multiple maturity dates, at each of which the policy may be "rolled over" for a further period; at that time if the guarantee is in the money, the insurer may pay out the difference; if it is out of the money, the guarantee may be reset at the higher current fund value; the policyholder may also have the option to reset at any time or at specified times for some minimum future period.

The methodology described allows for both deaths and withdrawals, and also for management charges and special charges for the guarantee. I sometimes feel that these practical complications, which of course must be allowed for by a real life office, serve to confuse the issue in a more theoretical exposition where one wishes to get over the fundamental principles. Fewer complications could have been included, but I do not feel that what is there is excessive.

However, although Professor Hardy shows how to obtain distributions of the costs, both on an emerging cash flow and on a present value basis, one thing that is missing here is how to calculated the charges, which is covered later in Chapter 11.

At this stage it is also assumed, without discussion, that the guarantee reserve is invested in risk-free instruments. This is probably the best strategy for this type of contract. But an alternative would have been to invest the reserves in the same fund as the policy. For other types of contract this might prove to be the better. It should be investigated too, as is done to some extent in WWY.

Chapter 7 is entitled "A review of option pricing theory" and it performs that function quite satisfactorily. As the author remarks, those who are familiar with the Black-Scholes principles can pass it by.

In Chapter 8 Professor Hardy explains how the option pricing methodology, with dynamic hedging, can be applied to the specific problem of investment guarantees. Although what is presented is quite clear, I would have taken it more slowly and more fully. Thus I would have started by demonstrating (for the benefit of the sceptics) that, if the real world behaves in accordance with the P-measure probabilities in the option pricing model, then dynamic hedging according to the Q-measure calculations does indeed provide investment proceeds that are close to what is required, and the more frequent the hedging the smaller the variance of the hedging error. Then one can go on, as the author does, to show that, even if the real world behaves according to some other model, in this case the RSLN model, then the proceeds may not be too far out, provided some of the parameters, in particular the variances, are comparable.

The very important point is made, which can hardly be over-emphasised, that one needs two models for these calculations, one an option pricing model which is used for calculating the option price and hedging quantities at each time step, and the other a model that simulates the real world in whatever way one wishes. I believe that lack of clarity about this may cause much confusion.

Professor Hardy sensibly shows a numerical example of dynamic hedging for a 2-year contract with no mortality and withdrawals, before going on to the complications of dealing with these decrements. I am a great believer in showing the simple case first. If it is confused with too may irrelevant features, the important points may be lost.

One aspect where I was not entirely happy with the explanation is in relation to calculating the present value of the "margin offset" charge. In Canada the guarantees are explicitly charged for by making a charge on the fund units every month of α times the amount of the fund at that time. This is in addition to, or part of, a management charge per month, and the total of them is m times the amount of the fund each month. Thus the invested fund increases at a rate less than the total return on shares (even if were to be invested in the share index). This in fact makes the guarantee more likely to be in the money at maturity. But to calculate the value of the margin offset, Professor Hardy sums the monthly charges, discounted at the risk free rate, and then takes the expected value under the Q measure. If we ignore the charges and other details one can get the present value, A, discounted at the risk free monthly rate r as:

$$A = E_Q[\Sigma_{t=0,n-1} \alpha.S_t e^{-rt}]$$

Where n is the number of months and S_t is the share index value at time t. A is then equated to the initial value of the option, B, to get a value for α .

It does not seem immediately clear why the Q measure is used, but I think it can be explained: we (the life office) wish to set up the hedging portfolio for the whole option initially. We require therefore to borrow an amount B. We can repay the loan from the future margin offset charges that we shall receive. The amounts of these will depend on the fund performance. But if we borrow shares of value B (or denominate the loan as if it were in shares), then, using shares as the numeraire, we do know what we shall receive, and we can repay the loan exactly as we receive the charges. This would justify discounting at the rate of return on the shares, and the result is certain, so we do not need to take expectations. We therefore put:

$$A = \sum_{t=0, n-1} \alpha. S_t S_0 / S_t = \alpha. S_0 [\Sigma_{t=0, n-1} 1]$$

And the answer, after allowing for the complications we have missed out, is the same as Professor Hardy gets. However, the process of financing the initial option value by borrowing shares is not explained. Effectively, the future margin offsets are hedged, which justifies using the Q measure, but the hedging is static, not dynamic, except that some of the loan is repaid every month.

An aspect where Professor Hardy treats things differently from the way WWY do is in the dynamic hedging process. Just before each rebalancing date

(at time t^-), the hedge portfolio has value $H(t^-)$; the desired value is H(t), and the hedging error is the difference between these. Professor Hardy assumes that the difference is made up at once (or taken away if it is a surplus), so that the investments at time t^+ are always what is required by the hedging process. She then discounts the hedging errors at the risk free rate to get a present value for them. This implicitly assumes that the hedging errors are financed by, or invested in, the risk free asset. WWY treat the affair differently. They assume that all that is available is $H(t^-)$, and they make alternative assumptions about how it is invested: (i) the right amount could be put into shares, with the balance invested in the risk-free asset; or (ii) the right amount could be put into the risk-free asset with the balance in shares; or (iii) the amount available could be invested in the right proportions. Option (i) is equivalent to what Professor Hardy has done, and it seems not unreasonable in this case that it turns out that the hedging error turns out to have lower variance under this option. But for other options the same result is not found. In my view one always needs to consider exactly how funds are invested or capital is financed, and not just assume that one should discount at any given rate.

In Chapter 9 risk measures are discussed, in particular quantile reserves (QR or VaR), and conditional tail expectations (CTE or Tail VaR). The latter have many desirable properties, and Professor Hardy, the Canadian Institute of Actuaries Taskforce and WWY all agree in preferring them to the former. Hardy shows how QR and CTE are related, how in some simple cases they can be calculated analytically, and how confidence intervals can be derived when they are simulated, all with practical examples. One nice feature is that graphs of distribution functions are drawn with the axes transposed (0 to 1 on the x axis, amounts on the y axis) so that a "more risky" distribution appears higher than a less risky one.

Hardy compares the QRs and CTEs found from the static (actuarial) and the dynamic (hedging) approaches, and shows that the latter gives (in her examples) lower extreme quantiles than the former, though the average cost/claim is often higher. This agrees with most of the results in WWY for GAOs, but they found that in some cases hedging gave even higher quantiles than the static approach.

A point not mentioned by Hardy is that CTEs allow an easy method of assessing the costs for individual policyholders as opposed to the costs for the whole portfolio; this is discussed in WWY. But a further point is that, although the CTE is analogous to a stop-loss calculation, being enough to provide a quantile reserves and also pay a "premium" for the average claim in excess of the QR, such insurance could not possibly be obtained at that price, so in effect the CTE, without reinsurance, is just a QR with a higher security level, a higher value of α .

In Chapter 10 the contracts are investigated using emerging cash flow analysis and profit testing, taking capital requirements into account. The distribution of profit using some desired rate of return on the capital required is the focus of interest. This is quite similar to what WWY have done, though the way that it is expressed by the different authors does not make this immediately clear. Hardy assumes the charge as given and calculates the expected profit and distribution of profit at different desired rates of return (risk discount rates).

WWY choose specimen rates of return, and calculate the break-even charge that results. But in both cases it is recognised that prudential reserves, whether these are part of the policy reserves or treated as solvency capital, are required, and the policyholders need to pay the average cost of their benefits, plus a "rent" for the use of this capital. So the premium they pay for the guarantee needs to be enough to cover both parts.

Hardy discusses the development of the prudential reserves (on a 95% QR basis) for specimen simulations, but does not bring out the additional aspect that the "fair value" of the contract, the price at which it could be transferred to another provider, which is what modern accounting principles are working towards, should be calculated on the same principles as the initial premium, as the expected value of the benefits (the "best estimate" perhaps) together with a sum that allows an adequate profit on the required contingency reserves. The fair value does not include the contingency reserves, but prudent reserves do include it. This is discussed by WWY, but whether the prudent reserves are part of the policy reserves or are part of the solvency capital, which in some countries may be of considerable practical significance, e.g. in relation to tax, is not considered, though Hardy mentions this point.

Chapter 11 discusses the important topic of forecast uncertainty (I should not say "important"; all the chapters in this book cover important topics). Professor Hardy attacks this in four steps: first, the errors from the random sampling inherent in Monte Carlo simulation; then variance reduction techniques; then parameter uncertainty; and finally model uncertainty.

Increasing the number of simulations reduces the random sampling errors, and it is useful to try out the convergence when the asymptotic result can be calculated analytically. The number of simulations required depends on the quantity we are estimating; tail values require more simulations than do means. A number of instructive examples are given.

Variance reduction techniques are also discussed, but Professor Hardy concludes that the only one that helps in this context is the control variate method. I had found, long ago, that some variance reduction techniques, such as importance sampling, were more trouble than they were worth, and indeed were sometimes so much slower than the simple method of just increasing the number of simulations was the best technique. The speed of computers has made it easier to do many more simulations. But one small feature that I discovered recently was that to calculate QRs or CTEs one needs to sort the results into order; many sorting routines increase in speed with the square of the number of cases sorted, and I found that the sorting took longer than the simulations had done; further investigation showed that a modern sorting technique (in fact Quiksort) improved my sorting speed over 100-fold, and that to take account of the fact that very many simulations gave guarantee costs of zero improved my sorting speed another 50-fold. Good computer algorithms, and also the source language one uses (complied or interpretive), can still make an enormous difference to computer run times. Looking carefully at your programmes may be a lot better than any variance reduction techniques.

Parameter uncertainty can be dealt with in three ways, of which Professor Hardy discusses only two, the Bayesian MCMC approach, and "stress testing"

by using alternative, but perhaps arbitrary, sets of parameters. With a complicated model it is not always easy to see which way one should move the parameters to test for stress, so I favour the "hypermodel" approach, by which I mean choosing, for each simulation, a random set of parameter values from some multivariate distribution for the parameters. Hardy uses the results from the MCMC approach; the alternative is to use the information matrix from the MLE method and to assume that the parameters are multivariate normally distributed. As I have noted above, this may require a careful choice of which parameter one chooses; $\log \sigma^2$ may be better than just σ . The multivariate normal method requires much less storage than MCMC, and it has the advantage that one can more easily tinker with the hyperparameters (the parameters of the distribution of the simulation parameters), and even splice together estimates from different investigations, which I suppose cannot be done with MCMC.

Model uncertainty is the last topic in this Chapter. Hardy's method is to try out alternative models. I would do just the same.

This ends the book's discussion of performance guarantees. Chapters 12 and 13 discuss, rather briefly, two extra topics: guaranteed annuity options (GAOs), and equity-indexed annuities. It is useful that these are mentioned, but a pity that they could not be fully developed. The paper on GAOs by WWY has 129 pages, and Yang's (2001) thesis is a great deal longer. Hardy gives 16 pages to the topic.

Some aspects of GAOs are similar to equity linked life insurance, in that, like them, the benefit can be defined as Max(A, B). Large contingency reserves may be required, and the actuarial and the hedging approaches are both possible. But GAOs have many different features. The type discussed by Hardy and by WWY is an equity-linked policy with a GAO at a fixed maturity (retirement) date, but in practice many of the policies issued in the U.K. have been with profits policies, with a range of possible maturity dates. Sometimes the guarantee is simply that a minimum amount of annual annuity will be available, rather than that the fund proceeds can be converted at a guaranteed rate; this is of course much cheaper.

To value GAOs one needs a stochastic model for interest rates, as well as for shares. A full yield curve model would be desirable, but one can do a lot with a model that allows for a level yield curve. Hardy has investigated U.K. data, and suggests two regime switching models, one for the FTSE All-Share index, one (with two autoregressive models) for the (long-term) yields on $2\frac{1}{2}\%$ Consolidated Stock ("Consols"), which is in effect a perpetual (and very old) British Government stock. This section is new material. But it is then shown how the actuarial approach can be applied, assuming that future mortality rates are known.

In practice future mortality rates cannot be forecast with certainty, and Yang (2001) investigates the effect of assuming a stochastic model (or "hypermodel") for forecast mortality rates. This is not just a matter of allowing for the random deaths in a small population of annuitants, but of allowing for the uncertainty of the underlying rates. Yang's method resembles that of Lee & Carter (1992), with some simplifications and some additional features. WWY show that the improvements in mortality in the U. K. since 1985 have been just

as important in increasing the cost of GAOs as the falls in interest rates that have occurred. Hardy does not discuss these points.

A further feature of GAOs is that, with a fixed guaranteed rate (Hardy, as Yang and WWY, uses £111 per annum per £1,000, though the actual rates offered by different offices vary considerably) the cost of the guarantee, however measured, varies very much with current interest rates, i.e. how much into or out of the money the guarantee is, whereas (at least under the lognormal model) the value of the equity linked investment guarantee is the same at all starting dates. This means that the uniform monthly charge, useable for equity linked guarantees, is unsuitable here, and an up-front charge, or at least a periodic charge that is fixed in advance and depends on the conditions at commencement is desirable.

GAOs lend themselves to option pricing models. It is convenient, though less realistic, to model the market annuity rate (at age 65) as a lognormal model, as do Yang and WWY. Full yield curve models after retirement have also been proposed, by e.g. Boyle and Hardy (2002). GOAs can be treated like a portfolio of bond "swaptions", as shown by Pelsser (2002). But an extra feature of the equity-linked GAO is that the amount to be converted depends on share price performance, so the option is analogous to a "quanto" option. This makes the option pricing mathematics harder to develop, but the results are not too difficult to understand. However, to hedge one needs to hold the full value of the policy including the option in shares, and then have offsetting long and short amounts, long in a portfolio that would replicate the deferred annuity and short in a zero-coupon bond maturing at the maturity date. The required amounts are the larger the more the option is "in the money". But whether long enough bonds to match the deferred annuity, and whether it is practicable to have large short holdings in zero-coupon bonds (unless they can be "borrowed" from the rest of the life office) are both doubtful.

Thus the dynamic hedging approach for GAOs may be impractical. It is therefore necessary for life office to consider the required contingency reserves, with both the static and the dynamic approaches. Hardy covers the main aspects well, but necessarily leaves much unsaid.

Equity-indexed annuities, covered in Chapter 13, are much simpler, appearing very similar to the equity-linked guarantee, but typically funded as an investment in bonds plus a call option, rather than as an investment in shares plus a put option. The term is typically much shorter, the option risk is often reassured with a third party, and the guarantee depends usually on the share price index, not a total return index. However, there are many interesting features of these contracts, including annual minima and maxima, and the possibility that the share return guaranteed is taken as only a fraction of the actual return. However as Hardy says, these policies are usually tackled as (possibly complicated) option pricing problems, and the actuarial method is normally absent.

This review is rather longer than is usual in ASTIN Bulletin, but I have had a lot to say on the subject. But modern reviews do not begin to compete with those of the 19th Century. Macaulay's review in The Edinburgh Review of Gleig's Memoirs of the life of Warren Hastings (1841) takes 140 pages in my

448 BOOK REVIEWS

reprinted (1898) copy. Macaulay's review is perhaps more worth reading nowadays than the book he was reviewing. This is not the case for this article. Read Mary Hardy's excellent book.

DAVID WILKIE

REFERENCES

- BOYLE, P.P. and HARDY, M.R. (2002) *Guaranteed annuity options*. Working Paper, University of Waterloo.
- HARRIS, G.R. (1999) Markov chain Monte Carlo estimation of regime switching vector autoregressions. ASTIN Bulletin 29, 47-80.
- LEE, R.D. and CARTER, L. (1992) Modelling and forecasting U.S. mortality, *Journal of the American Statistical Association* 87, 659-671.
- MATURITY GUARANTEES WORKING PARTY (1980) Report of the Maturity Guarantees Working Party. *Journal of the Institute of Actuaries* **107**, 103-209.
- Pelsser, A. (2002) Pricing and hedging guaranteed annuity options via static option replication. Working paper.
- WHITTEN, S.P. and THOMAS, R.G. (1999) A non-linear stochastic asset model for actuarial use. *British Actuarial Journal* **5**, 919-953.
- WILKIE, A.D., WATERS, H.R. and YANG, S. (2003) Reserving, pricing and hedging of policies with guaranteed annuity options. *British Actuarial Journal* 9, forthcoming.
- YANG, S. (2001) Reserving, pricing and hedging for guaranteed annuity options, PhD Thesis, Heriot-Watt University, Edinburgh.