

VALIDATION OF LONG-TERM EQUITY RETURN MODELS FOR EQUITY-LINKED GUARANTEES

Mary R. Hardy,* R. Keith Freeland,[†] and Matthew C. Till[‡]

ABSTRACT

A number of models have been proposed for the equity return process for equity-linked guarantees, following the introduction of stochastic modeling requirements by the Canadian Institute of Actuaries and the American Academy of Actuaries. In this paper we present some of the models that have become well known and discuss the use of residuals to test the fit. After showing that the use of the static, “actuarial approach” to risk management can result in two models with very similar likelihood and residuals giving very different capital requirements, we propose an extension of the bootstrap to compare all the models and to determine whether the optimistic or pessimistic view of the long-term left tail risk is more consistent with the data. Our context is the determination of capital requirements, so we are concerned in this work with real-world rather than risk-neutral processes.

1. INTRODUCTION

In 2001 the Canadian Institute of Actuaries Task Force on Segregated Funds (CIA 2001) proposed a stochastic methodology for the calculation of reserves and capital requirements for Segregated Fund contracts. In 2005 the American Academy of Actuaries followed suit with its C3 Phase 2 report (AAA 2005) for variable annuity contracts.

The stochastic valuation of equity-linked insurance, including segregated funds and variable annuities, requires a model for equity returns, which is used to generate a distribution for the future liability under the contract. In both the CIA and AAA reports, the precise nature of the equity return model was not mandated. Instead, actuaries are given the freedom to use any model, provided it can be calibrated to give distribution tails that are sufficiently fat—where “sufficiently” is defined in the calibration requirements of the respective reports.

At the time that the CIA task force began its deliberations, very few models were in common use. The Lognormal model, the Wilkie model (Wilkie 1995), and an empirical approach were cited in a survey. One company used a log-stable model.

Since the CIA report was published, a plethora of models has been proposed for segregated fund and variable annuity risk management purposes. We show in this paper that different models that look very similar may give very different results in practice. The objective of this paper is to identify why the differences occur and to look more closely at the fit of these models to the historical data.

It should be emphasized that we are dealing entirely in the realm of P -measure, the real-world measure, and that we are concerned with discrete-time models, which are the basis for capital requirement calculations for equity-linked insurance. P -measure is important for unhedged risk. Such risk arises because of a deliberate strategy not to hedge fully, or because in practice full self-funding hedging is impossible because of discrete hedging costs and other frictions. P -measure is the appropriate measure

* Mary Hardy, PhD, FIA, FSA, is CIBC Professor of Financial Risk Management in the Statistics and Actuarial Science Department, University of Waterloo, Ontario, Canada.

[†] Keith Freeland, PhD, ASA, is an Assistant Professor in the Statistics and Actuarial Science Department, University of Waterloo, Ontario, Canada.

[‡] Matthew Till is a graduate student in the Statistics and Actuarial Science Department, University of Waterloo, Ontario, Canada.

for calculating risk measures such as Value-at-Risk or the conditional tail expectation for capital requirements. In contrast, the risk-neutral measure, or Q -measure, is an artificial measure used to determine hedge costs. The econometrics of Q -measure requires matching market prices of options to a risk-neutral martingale measure. Some exploration of fitting and testing Q -measure models can be found in the work of Bakshi, Cao, and Chen (1997, 2000). We are not concerned with Q -measure here. The processes we discuss, with the maximum likelihood parameters, are not risk neutral, and we are not proposing their use for pricing options. However, they may be used to assess the adequacy of a risk mitigation strategy, as we have done in Section 5.

The econometrics of P -measure may be addressed by fitting a stochastic process to the history of the relevant time series. This is the topic of this paper. We are interested specifically in identifying which model shows the most consistency with the historical evidence, both in terms of the static discrete one-period fit and in the subsequent distributions of accumulation factors, which are products of the month-to-month returns (or sums of the log-returns), and which therefore require consistency with the longer-term relationships between consecutive returns. There are other methods for selecting models for economic capital, such as adherence to specific economic theories, but we are interested here in the narrow, but interesting, topic of goodness-of-fit to the available data. All the models discussed in detail in this paper are justified by their proponents in terms of their fit to the historical data.

We have emphasized the long-term nature of these models in the paper title. In selecting data to fit the models, we have chosen a fairly long history. Using a long history leads to different model selection decisions than using short terms. Consider only the past five years or so, and many models, both thin and fat tailed, will offer very similar fit statistics. Consider the past 40 years, and many models cannot match the tail experience.

Similarly, looking ahead for applications, for short-term problems the impact of using different models may be fairly minor. The difference for a one-year variable annuity (VA)-type product between using a regime-switching lognormal model and a regular lognormal model may be very modest. For an out-of-the-money VA with a long term, however, it is paramount that models generate adequately fat-tailed accumulation factor distributions.

2. EQUITY RETURN MODELS FOR EMBEDDED GUARANTEES

In this section we define and briefly discuss the major features of the various models proposed. In all of the following definitions, we let S_t denote the value of an equity index at time t , where S_0 is arbitrarily set at 1.00. As it is often convenient to model the log-return on equities, let Y_t denote

$$Y_t = \log(S_t/S_{t-1}).$$

Then, obviously,

$$S_t = S_{t-1} \exp(Y_t) = S_0 \exp(Y_1 + Y_2 + \dots + Y_t).$$

The n -period accumulation factor describes the growth of \$1 over n time units and is equal to $\exp(Y_1 + Y_2 + \dots + Y_n)$. In this paper we have used monthly time steps throughout, and all the models considered have been fitted by maximum likelihood to the S&P total returns data from January 1956 to September 1994.

2.1 The Independent Lognormal Model (ILN)

Under the independent lognormal model, the log-returns Y_t are assumed to have identical Normal distributions and are independent for $t = 1, 2, \dots$, that is,

$$Y_t = \mu + \sigma \varepsilon_t \quad \text{where } \varepsilon_t \text{ are i.i.d., and } \varepsilon_t \sim N(0, 1) \quad \forall t,$$

which means that, given only that $S_0 = 1$,

$$S_k \sim \log N(k\mu, k\sigma^2) \quad \text{for } k = 1, 2, \dots$$

This process requires only two parameters, μ and σ . The maximum likelihood parameter estimates, for the monthly S&P data from 1956 to 2004, are $\mu = 0.00834$ and $\sigma = 0.04234$.

The ILN model is the discretely observed version of geometric Brownian motion, which is one of the original assumptions of the Black-Scholes framework (Black and Scholes 1973). Short-term variations in equity returns often appear consistent with the lognormal model, which is why the basic Black-Scholes framework has proved remarkably robust for short-term financial instruments. However, as we will demonstrate, over longer terms the ILN model is generally far thinner tailed than the data, which matters for out-of-the-money options. All of the models we discuss below are fatter tailed than the ILN model.

In practice it is observed (for example, through implied volatility curves of traded options) that stochastic volatility models are required to model equity returns adequately. By stochastic volatility we mean any model for which the standard deviation of the outcome one or more time units ahead is modeled stochastically.

2.2 The GARCH(1,1) Model

The ARCH/GARCH family of models is discussed extensively in Engle (1995) and the references therein. The GARCH(1,1) model is commonly used to represent equity returns in option pricing (e.g., in Duan 1995). The autoregressive conditionally heteroscedastic family of models allow for stochastic volatility more than one period ahead, but, given the full history of the process, the one-period volatility is assumed fixed and known. In particular, the GARCH(1,1) model can be written as

$$\begin{aligned} Y_t | \mathcal{F}_{t-1} &= \mu + \sigma_t \varepsilon_t \quad \text{where } \varepsilon_t \text{ are i.i.d., and } \varepsilon_t \sim N(0, 1) \quad \forall t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (Y_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2. \end{aligned} \quad (2.1)$$

The standard filtration \mathcal{F}_{t-1} can be thought of as the relevant information for the process up to time $t - 1$. There is only one random process here, represented by ε_t . Also, given all the information at t (that is, Y_t and σ_t^2), the volatility at $t + 1$ is known. The volatility at any time from $t + 2$ forwards is random through the Y_t process.

If $\alpha_1 + \beta < 1$, then the process is covariance stationary, and we generally require this. In total the process depends on four parameters, $\{\mu, \alpha_0, \alpha_1, \beta\}$. Maximum likelihood is relatively straightforward for the conditionally heteroscedastic processes. Solver in Excel does the job quite efficiently. The maximum likelihood estimates for these parameters are $\mu = 0.00839$, $\alpha_0 = 0.00011$, $\alpha_1 = 0.0849$, and $\beta = 0.8563$.

The GARCH(1,1) model is a popular model for introducing stochastic volatility in a reasonably intuitive and tractable fashion (see, e.g., the seminal text by Campbell, Lo, and MacKinlay 1997). When assessing goodness-of-fit over medium terms (around five to 10 years, say), the GARCH model often provides a good parsimonious fit to the historical data. However, the GARCH model cannot capture large market crashes as well as some of the models in the regime-switching framework, described below.

2.3 The MARCH Family

Wong and Chan (2005) argue in favor of a mixture of ARCH models, which they call the MARCH family. The $\text{MARCH}(K; a_1, \dots, a_K; b_1, \dots, b_K)$ specifies a mixture of K ARCH models, where a_j is the autoregressive order of the j th model, and b_j is the ARCH order of the j th model. Wong and Chan specifically identify the $\text{MARCH}(2;0,0;2,0)$ model for log-returns, so that the model is a mixture of an ARCH(2) model and a random walk model, that is,

$$Y_t | \mathcal{F}_{t-1} = \begin{cases} Q_1 & \text{w.p. } q \\ Q_2 & \text{w.p. } (1 - q) \end{cases} \quad (2.2)$$

where

$$Q_1|\mathcal{F}_{t-1} = \mu_1 + \sigma_t z_t \quad \text{where } z_t \text{ are i.i.d., and } z_t \sim N(0, 1) \quad \forall t$$

$$\sigma_t^2 = \alpha_{1,0} + \alpha_{1,1} (Y_{t-1} - \mu_1)^2 + \alpha_{1,2} (Y_{t-2} - \mu_1)^2$$

$$Q_2|\mathcal{F}_{t-1} = \mu_2 + \alpha_{2,0} z_t.$$

Wong and Chan propose the model on the grounds that it offers an improved fit to the higher moments of the historical data, compared with the Regime-Switching Log Normal.

Note that there are two random processes involved: the mixture random variable and the normal innovation. There are seven parameters, $\{\mu_1, \mu_2, \alpha_{1,0}, \alpha_{1,1}, \alpha_{1,2}, \alpha_{2,0}, q\}$. The maximum likelihood estimates of the parameters are $\mu_1 = 0.012$, $\mu_2 = -0.0258$, $\alpha_{1,0} = 0.0011$, $\alpha_{1,1} = 0.0749$, $\alpha_{1,2} = 0.0802$, $\alpha_{2,0} = 0.004614$, $q = 0.910$. Like the regime-switching model, the MARCH describes two possible outcomes for each time unit; Q_1 represents a relatively low-variance process, with mean log-return 1.2% per month, but 9% of the values randomly come from a much more volatile regime, Q_2 , with a mean log-return of -2.58% , and a volatility of 6.8% (from $\sqrt{\alpha_{2,0}}$) per month.

2.4 The Regime-Switching Lognormal Model (RSLN)

The regime-switching lognormal model was proposed for the purpose of modeling long-term equity returns in Hardy (2001), with further discussion in Hardy (2003). The regime-switching framework was first introduced by Hamilton (1989). The essence of the model is that the process switches between two or more standard log-normal processes; the switching mechanism is another random process, generally assumed to be Markov: that is, the probability of switching regimes depends only on the current regime, not on the history of the switching process.

We find that for the data we consider (monthly stock index data) a two-regime lognormal model works best. The regime process is denoted by ρ_t , which takes the value 1 for the first regime, and 2 for the second. The log-return process can then be summarized as

$$Y_t|\rho_t = \mu_{\rho_t} + \sigma_{\rho_t} z_t,$$

$$\rho_t|\rho_{t-1} = \begin{cases} 1 & \text{w.p. } p_{\rho_{t-1},1} \\ 2 & \text{w.p. } p_{\rho_{t-1},2} = (1 - p_{\rho_{t-1},1}). \end{cases}$$

There are six parameters, $\{\mu_1, \mu_2, \sigma_1, \sigma_2, p_{1,2}, p_{2,1}\}$. The maximum likelihood estimates for the 1956–2004 S&P data are $\mu_1 = 0.0126$, $\mu_2 = -0.0097$, $\sigma_1 = 0.0342$, $\sigma_2 = 0.0635$, $p_{1,2} = 0.0432$, $p_{2,1} = 0.1834$. These are found using the method described in Hardy (2003), again using Excel Solver.

As with the MARCH model, there are two random processes. However, there is a difference between the mixture process and the switching process; under the mixture model, the probability of using either the Q_1 or Q_2 distributions is the same each time period. Using regime switching, the probability depends on which model was used in the previous time period.

The RSLN model with two regimes was one of the standard models used by the Canadian Institute of Actuaries Task Force on segregated fund contracts (CIA 2001). The volatility in this model follows a very simple two-state process, but because we are also changing the μ parameter at the same time, we can capture the association of high volatility with poor returns—that is, markets are more likely to fall steeply than increase steeply, when moving from a period of relative stability to a period of instability. The model is very tractable; quantiles and other risk measures can be calculated explicitly. The model thus offers a relatively straightforward way of capturing market behavior in a process that is even more tractable than the GARCH, though with more parameters and with two stochastic factors rather than one.

2.5 The Regime-Switching Draw-Down Model (RSDD)

Panneton (2002) proposes a variation on the regime-switching lognormal model. The regime-switching draw-down model adds a form of autoregression that influences the process when the index S_t falls below a previous high.

The definition of the RSDD process, for the two-regime version, is

$$Y_t | \rho_t = \kappa_{\rho_t} + \varphi_{\rho_t} D_{t-1} + \sigma_{\rho_t} \varepsilon_t,$$

where

$$D_{t-1} = \min(0, D_{t-2} + Y_{t-1}),$$

and ρ_t is defined as for the RSLN model above.

The D_t process tracks how far the total log returns have fallen below the previous high. The draw-down factor is the only difference between the RSLN and RSDD models, so the RSDD requires eight parameters, $\{\kappa_1, \kappa_2, \varphi_1, \varphi_2, \sigma_1, \sigma_2, p_{1,2}, p_{2,1}\}$. The maximum likelihood estimates are found very similarly to the RSLN model. They are $\kappa_1 = 0.0095$, $\kappa_2 = -0.0498$, $\varphi_1 = -0.0671$, $\varphi_2 = -0.1137$, $\sigma_1 = 0.0354$, $\sigma_2 = 0.0617$, $p_{1,2} = 0.0212$, $p_{2,1} = 0.1524$. It is interesting to note that although the estimates for the two volatility parameters are very similar to the RSLN estimates, the transition probability from the first to the second regime is rather smaller.

Like the RSLN model, the RSDD process involves two separate random processes. If the φ parameters are set to zero, we recover the RSLN model. In rising markets also, the process follows essentially the RSLN model, as the D_t process will be zero. However, the RSDD model is not at all tractable.

The parameters φ_j are generally negative. The intuition that is being incorporated here is that when returns are far below the previous high, the market exerts pressure for the index to recover. This appears a relatively minor adjustment to the RSLN model, and in comparing the distribution of monthly returns the models are practically indistinguishable. However, the effect on long-term returns is dramatic as the additional term causes a significantly thinner left tail for longer accumulation factors than the RSLN model. There are open questions here about the effect of survivorship—that one way for the recovery of an index such as the S&P 500 is the replacement of failing stocks with healthier alternatives, giving the index a better return than would be achieved by an investor.

2.6 The Regime-Switching GARCH Model

Combining the popular GARCH and RSLN frameworks, this model was described by Gray (1996) who describes a model that combines the main features of the GARCH model within a regime-switching framework. Let $p_1(t)$ denote the probability that the process is in regime 1 at t . Then

$$Y_t | \rho_t = \mu_{\rho_t} + \sigma_{\rho_t} \varepsilon_t,$$

$$\sigma_{\rho_t}^2 = \alpha_{\rho_t,0} + \alpha_{\rho_t,1} \varepsilon_{t-1}^2 + \beta_{\rho_t} \sigma_{t-1}^2,$$

where

$$\varepsilon_t = Y_t - (p_1(t)\mu_1 + (1 - p_1(t))\mu_2),$$

$$\sigma_t^2 = p_1(t)(\mu_1^2 + \sigma_{1,t}^2) + (1 - p_1(t))(\mu_2^2 + \sigma_{2,t}^2) - (p_1(t)\mu_1 + (1 - p_1(t))\mu_2)^2,$$

$$\rho_t | \rho_{t-1} = \begin{cases} 1 & \text{w.p. } p_{\rho_{t-1},1} \\ 2 & \text{w.p. } p_{\rho_{t-1},2} = (1 - p_{\rho_{t-1},1}). \end{cases}$$

This process involves 10 parameters: $\{\mu_j, \alpha_{j,0}, \alpha_{j,1}, \beta_j, p_{j,1}\}$ for $j = 1, 2$.

We recover the RSLN model when $\alpha_{j,1}, \beta_j$ are set to zero. The RSGARCH model is significantly more complicated than the RSLN or GARCH models. It is included here because it combines some of the advantages of each individual model; like the RSLN model, the RSGARCH incorporates the association of crashes with poor returns, which is not captured by the GARCH model. However, like the GARCH

model, the RSGARCH model allows a wide range of possible values for the volatility, unlike the RSLN model, which (in the two-regime version) only permits the volatility to take one of two values.

Estimation for this model can be problematic as the likelihood surface has many local maxima, many of which are very close to the global maximum. This means that many parameter sets are more or less equally valid under maximum likelihood estimation, even though the outcome may look very different. This means that there is very large uncertainty about the parameters, particularly those associated with the high-volatility regime. The maximum likelihood parameters used are $\mu_1 = 0.0105$, $\mu_2 = -0.0592$, $\alpha_{1,0} = 0$, $\alpha_{2,0} = 0.0024$, $\alpha_{1,1} = 0$, $\alpha_{2,1} = 0$, $\beta_1 = 0.8508$, $\beta_2 = 0.8610$, $p_{12} = 0.0455$, $p_{21} = 0.9835$, but quite different parameters give likelihood values very close to maximum.

2.7 The AAA Stochastic Log-Volatility Model

The American Academy of Actuaries (AAA, 2005) in its C3 Phase 2 document proposes a “stochastic log-volatility” model. The basic model description is as follows. The mean and volatility are expressed as annual rates and then adapted for application to the monthly process. This is consistent with the description in the AAA document:

$$Y_t = \mu_t/12 + (\sigma_t/\sqrt{12}) z_{y,t},$$

where

$$\begin{aligned} \mu_t &= A + B\sigma_t + C\sigma_t^2, \\ \log \sigma_t = \nu_t &= (1 - \varphi)\nu_{t-1} + \varphi \log \tau + \sigma_v z_{\nu,t}, \end{aligned}$$

where $(z_{y,t}, z_{\nu,t})$ have a standard bivariate normal distribution, with correlation ρ . Upper and lower bounds to constrain the log-volatility process are omitted for clarity but have been incorporated in the estimation processes in the following sections. The lower bound for annual volatility is 3.05%, the upper bound is 79.88%, and the upper bound for $\exp((1 - \varphi)\nu_{t-1} + \varphi \log \tau)$ is 30%.

The AAA C3 phase 2 report claims that this model “captures the full benefits of stochastic volatility in an intuitive model suitable for real world projections.” The model is included here because it was used to set the calibration requirements in the C3 Phase 2 report and was used for prerecorded scenarios for plug-and-play use by insurers. The model has six parameters $(\varphi, \tau, A, B, C, \rho)$, as well as the bounds.

The SLV model is more complicated to estimate. Recall that the maximum log-likelihood is the maximum value over the parameter space of

$$\sum_{t=1}^n \log f(y_t | \Theta, y_1, y_2, \dots, y_{t-1}),$$

where Θ represents the parameter set, and f is the density function for the log-return. Under the SLV model we have a random log-volatility process,

$$\nu = \{\nu_1, \nu_2, \nu_3, \dots, \nu_n\}.$$

Given the log-volatility ν , the density function for $y_t | y_1, \dots, y_{t-1}, \Theta$ is simply

$$f(y_t | y_1, \dots, y_{t-1}, \Theta, \nu) = \frac{1}{\sigma_t/\sqrt{12}} \phi \left(\frac{y_t - \mu_t/12.0}{\sigma_t/\sqrt{12}} \right),$$

where $\sigma_t = e^{\nu_t}$.

The volatility process is too complex to calculate this directly, but we can use Monte Carlo simulation to estimate the expectation of $f(y_t | y_1, \dots, y_{t-1}, \Theta, \nu)$ with respect to ν : that is, generate N volatility paths, each with n values, where n is the length of the data series. Let ν_j denote the j th simulated log-volatility path, then the estimated contribution to the log-likelihood from the t th observation is

$$\sum_{j=1}^N \frac{\log f(y_t|y_1, \dots, y_{t-1}, \Theta, \nu_j)}{N}.$$

Shephard (2005) describes more sophisticated simulation procedures, using importance sampling, for general stochastic volatility models, but straightforward Monte Carlo appears to work well here, with very little sampling variability. Using only 1000 simulations, the standard error of the log-likelihood is around 0.2. The estimated maximum likelihood parameters are $\varphi = 0.35$, $\tau = 0.137$, $A = 0.055$, $B = 0.482$, $C = -0.9$, $\rho = -0.25$.

3. MAXIMUM LIKELIHOOD RESULTS

We have found the maximum likelihood parameters for each of the models and the corresponding value of the maximum log-likelihood based on the 584 values of the monthly S&P 500 total return index from January 1956 to September 2004.

The log-likelihoods are given in Table 1. Clearly the lognormal model is an outlier, with a very poor overall fit. The other six models appear comparable.

Likelihood-based model selection offers the Akaike Information Criterion, which says that each additional parameter should contribute at least one unit to the log-likelihood, which would suggest that the RSGARCH model is the preferred, followed by the RSDD, RSLN, MARCH, SLV, and GARCH.

The Schwarz-Bayes information criterion (BIC) is a little more sophisticated, as it takes the sample size into consideration. According to the BIC, each additional parameter should increase the log-likelihood by around 3.2 units. This would indicate that the RSGARCH is the best fit, followed by the RSLN and RSDD, the GARCH and MARCH, the SLV, and finally the ILN.

However, likelihood-based selection does not address the central questions: Do these models fit the data? Or, the more relevant question: Do these models fit the data in the parts of the distribution most critical for the equity-linked capital requirement calculations? How much does the fit quality matter in the final calculation? Do the models adequately fit the aggregated data, for the long-term accumulation factors?

4. RESIDUAL ANALYSIS

For each of these models there is a sequence of residuals (or combination of residuals for the regime-switching models) that should be, at least approximately, an i.i.d. normally distributed sample according to the assumptions of the model. For all the models that are not mixed or regime-switching, the residuals are exactly $N(0,1)$, under the null hypothesis. If we examine the residuals and they are far away from the Normal model, that's a signal that the fit is not adequate.

Table 1
Maximum Log-Likelihoods for the Seven Models

Model	Number of Parameters	Maximum Log-Likelihood
Lognormal	2	1018.0
GARCH(1,1)	4 (+1 starting volatility value)	1030.1
MARCH	7	1039.8
RSLN	6	1042.0
RSDD	8 (+1 starting "drop-down" factor)	1047.1
SLV	7 (+3 min & max constraints)	1035.2
RSGARCH	10 (+1 starting volatility value)	1056.0

For the MARCH and regime-switching processes, the residuals are only approximately $N(0,1)$, because of the uncertainty in the regime (or mixture) process—that is, even when we know the log-return, we cannot be sure which regime the process is in. If there were no regime uncertainty, the residuals would be exactly $N(0,1)$ under the null hypothesis. The nature of the data is that the uncertainty is not great most of the time. The probability of being in, say, the first regime is close to one or zero for most observations. For the mixture distributions we have explored two ways of determining the residuals, using the residuals conditional on the regime. The first is to weight the two residuals according to the conditional probability for each regime. The second is to use a zero-one weighting, where we use the residual with the higher associated probability.

In other words, using the RSLN as an example, we calculate

$$r_{t,1} = r_t | (\rho_t = 1) = \frac{y_t - \mu_1}{\sigma_1}, \tag{4.1}$$

$$r_{t,2} = r_t | (\rho_t = 2) = \frac{y_t - \mu_2}{\sigma_2}. \tag{4.2}$$

Then we use the conditional probability

$$p(1) = \Pr[\rho_t = 1 | \{y_j\}_{j \leq t}]$$

for the weighted average residual

$$r_t = p(1) r_{t,1} + (1 - p(1)) r_{t,2},$$

or we use an indicator function for the zero-one residual

$$r_t = I_{\{p(1) > 0.5\}} r_{t,1} + (1 - I_{\{p(1) > 0.5\}}) r_{t,2}. \tag{4.3}$$

Both the probability weighted and the zero-one weighted residuals give very similar results; the plots and figures quoted below use the zero-one weighting system.

We use two methods for exploring the residuals. The q - q plot shows the model residuals quantiles against the Normal distribution quantiles. If the model assumptions hold, the residuals should lie close to the central diagonal. In Figures 1 and 2 we show the q - q plots for all seven models. If the residuals are normal, they would lie on the diagonal shown.

The second item is the Jarque-Bera test of normality, a statistical test often used for residual analysis. It is a simultaneous test of the skewness and kurtosis of the residuals. The test statistic uses the residual skewness S and excess kurtosis K :

$$Q = \frac{n}{6} \left(S^2 + \frac{K^2}{4} \right).$$

Under the null hypothesis, that the residuals are normal, Q has a χ^2 distribution with two degrees of freedom. For every model we consider, we can examine the raw residuals or some combination of residuals that should be independent and, at least approximately, normally distributed, according to the original model specifications. The Jarque-Bera test is described in more detail in Jarque and Bera (1980, 1987). Failing this test indicates that the residuals are not normal, which is an indication that the model is not consistent with the data, or, for the MARCH and regime-switching models only, it could indicate that the effect of regime uncertainty is too large for the residuals to appear normal.

For all the residual analysis we have used maximum likelihood estimates of all parameters in order that the models can be directly compared.

4.1 ILN

For the fitted values of μ and σ , and the observed values of the log-returns, y_t , the ILN residuals are simply

Figure 1
q-q Plot of Residuals under ILN, GARCH, MARCH, and SLV Models

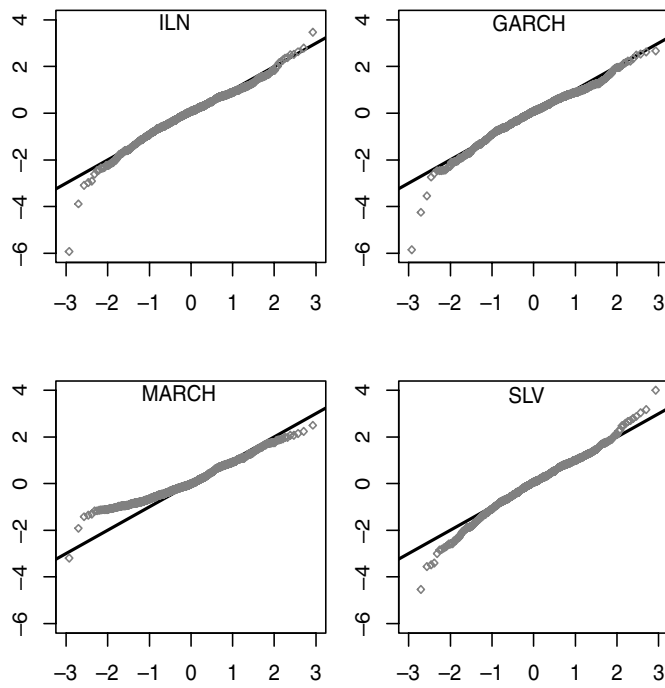
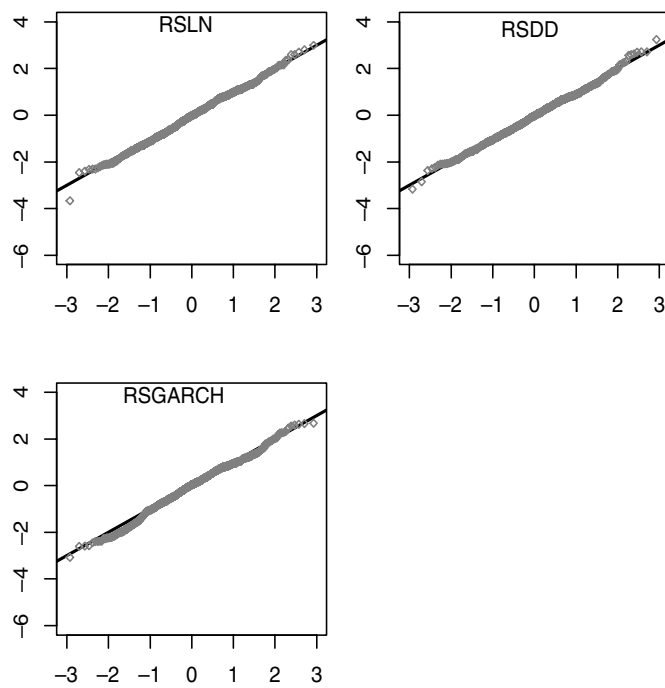


Figure 2
q-q Plot of Residuals under Regime-Switching Models



$$r_t = \frac{y_t - \hat{\mu}}{\hat{\sigma}},$$

which under the ILN assumptions are i.i.d. $N(0,1)$ distributed.

The q - q plot is shown in Figure 1. The fit fails, predominantly, in the left tail, where the residuals are much fatter tailed than the Normal. The failure of the fit on the left tail causes the failure of the Jarque-Bera test of normality, with a p -value of around 10^{-40} .

4.2 GARCH

The GARCH residuals are

$$r_t = \frac{y_t - \hat{\mu}}{\sigma_t},$$

where

$$\sigma_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1(y_{t-1} - \hat{\mu})^2 + \hat{\beta} \sigma_{t-1}^2.$$

The q - q plot is shown in Figure 1.

Again, the left tail fit is poor, and this is reflected in a Jarque-Bera test p -value, once again, around 10^{-40} .

4.3 MARCH

The MARCH is a mixture of two models, each with $N(0,1)$ i.i.d. residuals under the model assumption. We have treated this similarly to the regime-switching models (it is a special case of a regime-switching model with two regimes, where $p_{11} = p_{21} = q$), using the zero-one weighting for the two regimes. This means that we calculate the residuals conditional on the return coming from the Q_1 process, and separately assuming the Q_2 process. We calculate the probabilities that the return is from each process, conditional on the return value, and we use whichever return is most likely.

The q - q plot is given in Figure 1. The fit is better than the ILN, GARCH, and SLV models, but not as good as the regime-switching models, with a p -value for the Jarque-Bera test of 0.006. However, the residuals are only approximately normal under the null hypothesis, so this test is inconclusive. We cannot say whether the failure arises because the null hypothesis should be rejected, or because the approximation of normality for the mixed residuals is too inaccurate here.

4.4 SLV

The residuals for the SLV model depend on the random volatility process. Using the same technique as for the log-likelihood, we can estimate the residuals by Monte Carlo simulation of the log-volatility paths, and, as with the log-likelihood, the sampling variability is small. Using N simulations of the log-volatility path ν , such that $\sigma_{j,t}$ is the t th simulated volatility from the j th simulated path, the residuals are estimated from

$$\tilde{r}_t = \frac{1}{N} \sum_{j=1}^N \frac{y_t - \hat{\mu}_t/12}{\hat{\sigma}_{j,t}/\sqrt{12}}.$$

The q - q plot is given in the lower right plot of Figure 1. The fit is not very good in either tail, and the Jarque-Bera test fails with a p -value of 10^{-40} .

4.5 RSLN

The RSLN residuals calculation was described in equation (4.3). We have shown in Figure 2 the q - q plot of the zero-one weighted residuals. The residuals easily pass the Jarque-Bera test with a p -value of 0.8.

4.6 RSDD

The RSDD residuals were calculated similarly to the RSLN, with

$$(r_t | \rho_t = 1) = \frac{y_t - (\hat{\kappa}_1 + \hat{\phi}_1 d_{t-1})}{\hat{\sigma}_1},$$

where $d_{t-1} = \min(0, d_{t-2} + y_{t-1})$, and similarly for $r_t | \rho_t = 2$. We assume $d_0 = 0$. We then apply the zero-one weighting to the residuals.

The q - q plot shows the residuals are slightly closer to the normal quantiles than the RSLN model in the left tail. The residuals pass the Jarque-Bera test with a p -value of 0.73.

4.7 RSGARCH

Similarly to the other RSDD models, we calculate the residuals for the two regimes, for example,

$$(r_t | (\rho_t = 1)) = \frac{y_t - \hat{\mu}_1}{\sigma_1(t)},$$

and then use whichever regime is more likely based on the conditional probabilities. The results, in Figure 2, are similar to the other regime-switching models, and the residuals pass the Jarque-Bera test with a p -value of 0.90.

4.8 Summary

Only the regime-switching models pass the Jarque-Bera test, though as noted above, the test is approximate only for the mixture models and may in particular be misleading for the MARCH model. The reason for the failure of the non-regime-switching models is clear from the q - q plots, which show that all the others fail in the left tail of the residuals. The SLV model does not appear to offer a very good fit of the right tail either.

There is no way to tell from the residuals analysis whether any of the regime-switching models is better than the others.

5. HOW MUCH DOES IT MATTER?

All of these models produce comparable maximum likelihood values, indicating an acceptable overall fit. The residuals show that the models may be separated into two or three groups—the regime-switching models appear to offer a reasonable fit, the others, less so. The poor fit for, say, the GARCH model arises from a very small number of outlying data points; however, it is precisely these results that could be important for calculating tail measures for longer-term accumulation factors.

If there is little difference in the capital requirements generated using the different models, then it would not matter too much which is used in practice. In this section we apply all of the models of the previous section to a sample variable annuity-type contract. We do not calibrate the models to the AAA or CIA tables; instead we use the maximum likelihood parameter estimates in the scenario generation. The reason is that we are interested in whether the calibration requirements are justifiable, or whether any model producing similar maximum log-likelihood is, essentially, just as valid as any other.

We will look at the implications for capital requirements based on a traditional static actuarial approach and, separately, using a simple delta hedge.

5.1 Example Contract

To illustrate the effect of model uncertainty we use a 20-year single-premium guaranteed minimum accumulation benefit (GMAB) issued to a life age 50. Some of the assumptions are simplistic; the example is used solely for illustration. The single premium is invested in the policyholder's fund, F_t , which is assumed to be invested in the S&P 500 index, with dividends reinvested. The benefit on death

or maturity is the greater of the accumulated investment proceeds and a guarantee. The guarantee is set at issue equal to the amount of the single premium. After 10 years, if the policy is still in force, the guarantee is reset to the accumulated investment proceeds, if that is greater than the guarantee. If the guarantee exceeds the fund, then a cash payment equal to the difference is paid into the fund at that time.

More mathematically, we have a guarantee level G_t at t with initial level G_0 equal to 100% of the single premium invested. Then

$$G_t = \begin{cases} G_0 & 0 \leq t \leq 10 \\ \max(G_0, F_{10}) & 10 < t \leq 20 \end{cases} \quad (5.1)$$

Also, we have potential option payouts under the GMAB at times $T = 10$ and $T = 20$; In each case the payout at F_T is $\max(G_T - F_T, 0)$. For a more detailed, and more general, description of the GMAB, see Hardy (2003).

We assume the policyholder is age 50 at issue, and that mortality follows the rates given in the appendix of Hardy (2003). Lapses are assumed level at 8% per year (for simplicity). A management charge of 300 basis points (bps) (3%) per year is deducted from the policyholder’s fund, of which 20 basis points is used to fund the guarantee.

There are two common approaches to the risk management of segregated fund and variable annuity contracts, of which this is a simplified example. The “actuarial approach,” which may be described as a static partial hedge, uses Monte Carlo simulation to project the liabilities, using a real-world distribution (not risk-neutral), discounted at the risk-free rate (assumed here to be 5%) to give a distribution of the present value of the liabilities if no hedging strategy is adopted. We apply a risk measure to the distribution to ascertain the total economic capital required for the contract.

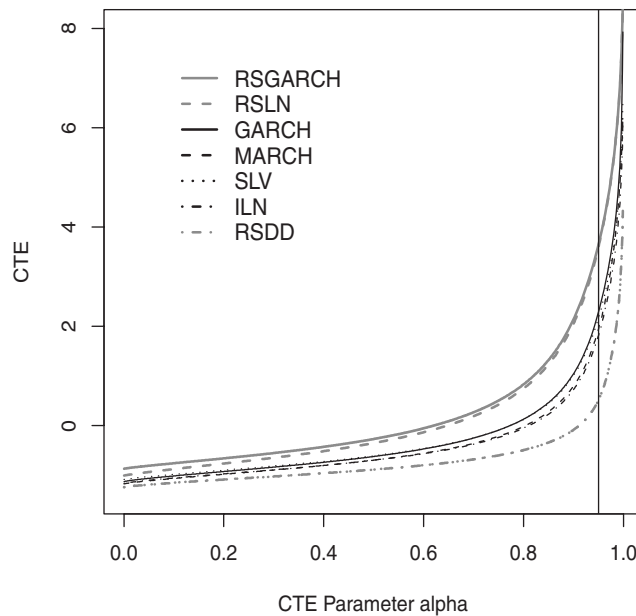
The other approach is to use dynamic hedging to limit the risk. The simplest version of this is to assume a simple Black-Scholes hedge is established at the contract inception, and rebalanced monthly. The hedge will not be perfect for a number of reasons, including discrete hedging error (a perfect hedge requires continuous rebalancing), transactions costs on the rebalancing of the hedge assets, and model error, because the log-normal model of Black-Scholes is, as we have seen, not a particularly good fit to the historical equity returns data. These hedge imperfections constitute an additional liability, which we can estimate by Monte Carlo simulation under a real-world measure. The process is described in more detail in Hardy (2003). We assume a risk-free rate of interest of 5% per year, and a volatility for hedging purposes of 18% per year, which is broadly consistent with the volatility in the models and data.

The risk measure used is the Conditional Tail Expectation (CTE), which is the basis of both the CIA segregated fund report (CIA 2001) and the AAA (2005) C3 Phase 2 report. The CTE at α , $0 \leq \alpha \leq 1$, is estimated as the average of the worst $100(1 - \alpha)\%$ of the simulations. For all the simulations summarized below we have used 10,000 projections, so the 95% CTE reported is the mean of the highest 500 simulated liability values.

In Figure 3 we show the results of using each model for the actuarial approach. The x -axis represents the CTE parameter. At $\alpha = 0$ we have the mean loss for the contract. As α moves near to one, the curve tends to the maximum of the simulated losses. The vertical line lies at $\alpha = 0.95$, corresponding to the standard used for the total balance sheet requirement for segregated funds. In Table 2 we show explicitly the 95% CTEs for the different models.

The graph shows that the results fall into three groups. The RSGARCH and RSLN models generate similar CTE results, with the highest values for the CTEs for all values of α . The next group contains the GARCH, MARCH, SLV, and ILN models. The 95% CTEs for this group are more than 1.5% less than the top group. At the bottom, offering the most optimistic view of the loss distribution is the RSDD model, which generates a 95% CTE more than \$3 smaller for every \$100 premium than the other regime-switching models. This difference could be very significant; suppose an insurer uses the RSDD model, but in fact the RSLN proves to be more accurate. For every contract on the cohort, the economic

Figure 3
CTEs Using Actuarial Approach, MLE Model Parameters



capital would be inadequate by 3% of the single premium, even after allowing for all margin offset income.

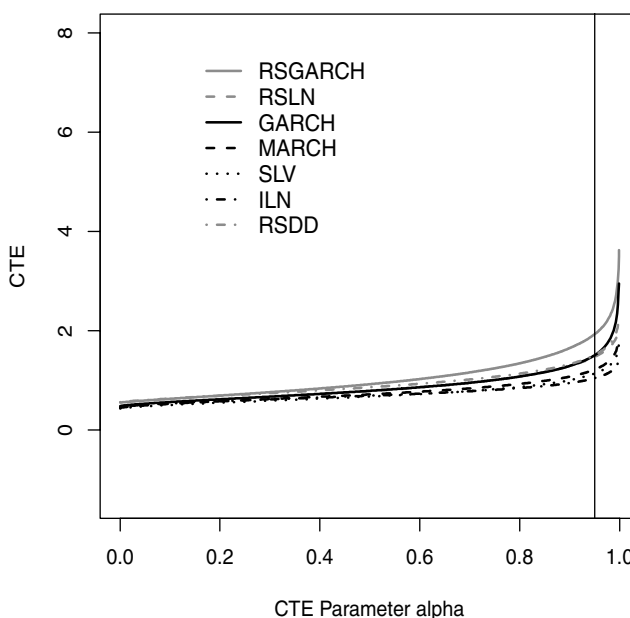
The split of the results, with the regime-switching models generating both the most optimistic and the most pessimistic views of the loss distribution, is somewhat surprising, as the analysis of likelihoods and residuals indicated that the regime-switching models offered a very similar fit to the data. Moreover, the CTEs using the RSDD model are substantially lower even than the ILN model, which offers a very poor fit to the data, in particular in the crucial left tail area of the returns distribution.

In Figure 4 we show the CTE results using the dynamic hedging approach, where the different models are used to project the hedge to estimate the distribution of the hedge costs and unhedged liabilities. The figure is plotted on the same scale as the actuarial approach in Figure 3. The 95% CTEs are given in Table 3. We notice here that the hedge clearly achieves a lot of risk mitigation—the maximum losses are much lower than under the actuarial approach, and the 95% CTE is lower using the dynamic hedging approach than it is under the actuarial approach for all models except the RSDD. The right tail protection is achieved at a cost, however, since the mean outcome is a loss using the hedging approach

Table 2
95% CTE for GMAB, Percentage of Single Premium, Actuarial Risk Management, MLE Model Parameters

Model	95% CTE
RSGARCH	3.64%
RSLN	3.59
GARCH	2.29
MARCH	1.94
SLV	2.16
ILN	1.81
RSDD	0.51

Figure 4
CTEs Using Dynamic Hedging Approach, MLE Model Parameters



compared with a profit under the actuarial approach (which indicates that the margin offset is too low, if the other assumptions are appropriate). The mean outcome is the CTE evaluated at $\alpha = 0$.

So it appears that model selection is not so critical using the dynamic hedging approach, but may have significant impact using the actuarial approach. An explanation is that the models all look fairly similar in the monthly returns. For the dynamic hedging approach, where the month-to-month equity index movement is the key factor, this means that the results are robust with respect to model variation. However, the actuarial approach depends on the long-term accumulations under the models. As we have seen, the RSDD model appears to behave very similarly to the RSLN (for example) on a month-to-month basis, but the differences over long accumulation periods are significant.

The actuarial approach to risk management is the dominant method in Canada and possibly also in the United States. It therefore matters whether the RSLN or the RSDD is a better model. If it is the RSDD and we assume the RSLN, we waste resources with unnecessary solvency capital. Vice versa, and we risk significant capital shortage in the event of poor market returns. In the following sections we will try to extract more information from the data using bootstrap techniques.

Table 3
95% CTE for GMAB, Percentage of Single Premium, Dynamic Hedging Risk Management, MLE Model Parameters

Model	95% CTE
RSGARCH	1.93%
RSLN	1.48
GARCH	1.51
MARCH	1.22
SLV	1.14
ILN	1.05
RSDD	1.51

6. BOOTSTRAPPING THE DATA

6.1 Bootstrapping Time Series

The bootstrap is a technique for exploring the relationship between a sample and an underlying population. The essential idea is that of resampling from the sample—that is, creating a new sample, the same size as the original, by drawing with replacement from the original sample. The relationship between the new, “pseudo-sample” and the original sample mirrors in many ways the relationship between the original sample and the underlying population. Hence, the bootstrap can offer a nonparametric (or parametric) estimate of the uncertainty of statistics such as quantiles estimated from samples. Efron and Tibshirani (1993) describe the basic techniques.

The application of bootstrap techniques to time series is a more recent development. With time series, simply drawing with replacement will not replicate the original distribution, as it will lose the serial dependency in the data. To retain dependency, blocks of data are drawn rather than individual values. The optimal block size is not always evident. If blocks are too short, dependency in the original sample will be lost. If the correlations are broadly positive, then the pseudo-sample will be too thin tailed, as the blocks will not adequately pick up periods of prolonged poor returns. If the time series is essentially serially independent, then short blocks will not affect the results. If the correlations are negative, then block sizes that are too short will cause the pseudo-sample to be too fat-tailed, as the original smoothing of results from consecutive observations will not be reflected in the pseudo-sample. If the block sizes are too large, we lose tail information in the data; the blocks effectively reduce the sample size. This will cause the sample to be thinner tailed than the original sample. After some exploration, we used six-month block sizes; the results were not sensitive to block sizes of between two and 12 months.

6.2 Accumulation Factors

The n -year accumulation factor is the accumulation after n years of an investment of \$1. In terms of the log-return random variable, Y_t , we are interested in the sums of consecutive values, which give the log of the accumulation factor. The CIA (2001) and AAA (2005) requires models to be calibrated to tables of accumulation factors for one, five, 10, and (for the AAA) 20 years. The CIA sets standards for the left 2.5%, 5%, and 10% quantiles. The AAA sets standards for both the left and right tails, up to the lower and upper 20% quantiles.

The index data we use cover 48 years. So for the one-year factor we have 48 nonoverlapping observations. We may consider the minimum value observed, which (using January to January data) is 0.770, as an estimate of the 1/49 or 2.04% quantile; the second smallest value, 0.835, is an estimate of the 2/49 or 4.08% quantile. Interpolating between these values gives an estimate of 0.7844 for the 2.5% quantile. Similarly, we have an estimate for the 5% quantile of 0.836, and for the 10% quantile we have 0.859.

In fact, even before we bootstrap, the data tell us more, since we get a slightly different story by looking at the years from February to February, March to March, etc. The different starting points make little difference to the 10% quantile estimate, but are more significant for the 5% and 2.5% quantiles. The range for the 2.5% estimate is (0.638,0.836).

Using standard bootstrap techniques we generate a 10,000 samples of 48 one-year accumulation factors by drawing with replacement in a four-month block. We calculate the estimate of the 2.5%, 5%, and 10% quantiles for each pseudo-sample just as we did for the original data. We can use the bootstrap sample to construct a confidence interval for the true quantiles. We compare this with the parametric estimates from the models. Any model that generates quantiles outside the bootstrap confidence interval is considered to have failed the test of fit.

The results for the one-year factors are shown in Table 4. We see that, even though the regime-switching models are fatter tailed than the other models, none of the models fail the bootstrap test.

We repeat the exercise for the 10-year accumulation factors. Here, with 48 years of data, we only have four nonoverlapping observations. The minimum value is an estimate of the 20% quantile, which

Table 4
Model and Bootstrap Quantiles for One-Year Accumulation Factors

Bootstrap 90% Confidence interval	2.5%-ile (0.67, 0.87)	5%-ile (0.76, 0.91)	10%-ile (0.84, 0.97)
Model:	2.5%-ile	5%-ile	10%-ile
ILN	0.829	0.868	0.916
GARCH	0.812	0.857	0.912
MARCH	0.823	0.868	0.918
SLV	0.825	0.868	0.915
RSLN	0.764	0.829	0.908
RSDD	0.768	0.831	0.901
RSGARCH	0.792	0.847	0.910

is as far into the tail as we can explore by standard methods. The empirical estimate ranges from around 1.2 to 2.2. The bootstrap 90% confidence interval and model values for the 20th percentiles are given in Table 5. Again, we find that none of the models generate quantiles outside the bootstrap interval, which is very wide.

7. OVERSAMPLING

The bootstrap, conventionally applied, did not help to identify whether any of the models are more consistent with the data. In this section we consider the effect of extending the time series bootstrap. For the 10-year accumulation factor, the bootstrap method requires the construction of sets of only four values. What are the implications of continuing to sample from the data and using the resulting distribution? That is, suppose we sample 10,000 10-year accumulation factors from the data, in blocks of six months. How would that distribution compare with the underlying population, on average?

The answer is that, if the data are essentially uncorrelated or positively correlated, in the sums as well as the individual months, oversampling will give a distribution that is thinner tailed, on average, than the original population.

The intuition is that by oversampling we are exploring regions of the distribution outside of the range of the data, on average. For example, suppose we have a random sample of four values from a Uniform(0,1) distribution. On average, the minimum value will be around 0.2. No matter how much we oversample from this set of four observations, we will not, on average, sample at all from the bottom 20% of the distribution. Thus, we will end up with a thinner-tailed distribution than the underlying population.

Table 5
Model and Bootstrap Quantiles for 10-Year Accumulation Factors

Bootstrap 90% Confidence interval	20%-ile (0.95, 2.83)
Model:	
ILN	1.841
GARCH	1.847
MARCH	1.909
SLV	1.788
RSLN	1.773
RSDD	1.953
RSGARCH	1.660

However, we know that for negatively correlated distributions, it is possible for blocking to fatten tails if the block sizes are not big enough. For positively correlated distributions, blocking with small blocks will thin the tails. So the dual effect of oversampling and blocking will result in thinner tails than the underlying population for zero and positively correlated data, but may combine for thinner or fatter tails, or neither, for negatively correlated data.

In Figure 5 we plot the partial autocorrelation function for both the original log-returns and for the sums of the log-returns. The graphs show that there is no evidence of negative correlation, over short- or long-term accumulations. Hence, we expect oversampling to generate a thinner tail for the accumulation factors than the underlying population.

We use this result by repeating the bootstrap test. This time, we estimate the left tail quantiles of the oversampled distribution for the 10-year accumulation factors. We compare these quantile estimates with those from the different models proposed. The oversampled distribution should be thinner tailed (on average) than the population. Thus, the left tail quantiles from the oversampled distribution should be larger than the underlying population. A model with smaller left tail quantile values than the oversampled distribution passes the test. Larger left tail quantiles imply a thinner tail than the oversampled distribution, and such a model is less likely to generate the data than the fatter tailed, and so fails the test. The results will apply for quantiles outside the range of the original data, which means that the test will be more useful for comparing the 10-year accumulation factors than for the

Figure 5

Partial Autocorrelation Functions for Monthly, Yearly, Five-Yearly, and 10-Yearly Accumulation Factors

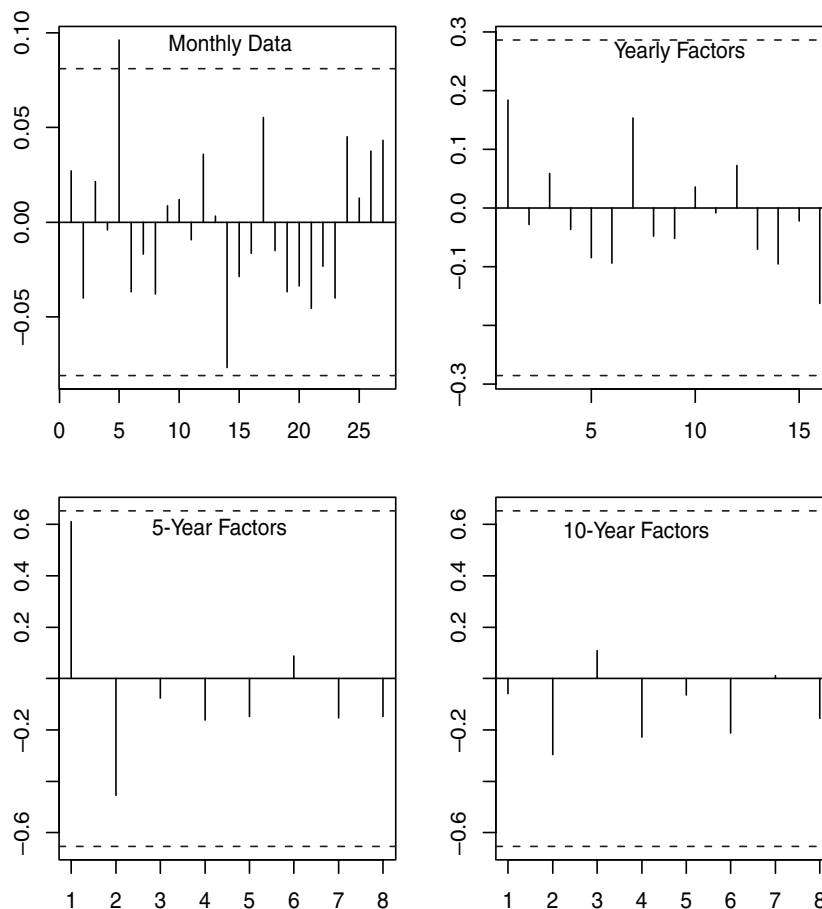


Table 6
Model and Oversample Quantiles for 10-Year Accumulation Factors

Model	Quantile		
	2.5%	5%	10%
Oversample distribution	1.041	1.228	1.478
ILN	1.096	1.269	1.501
GARCH	1.074	1.259	1.509
MARCH	1.131	1.297	1.558
SLV	1.087	1.249	1.467
RSLN	0.914	1.105	1.378
RSDD	1.277	1.439	1.653
RSGARCH	0.905	1.086	1.315

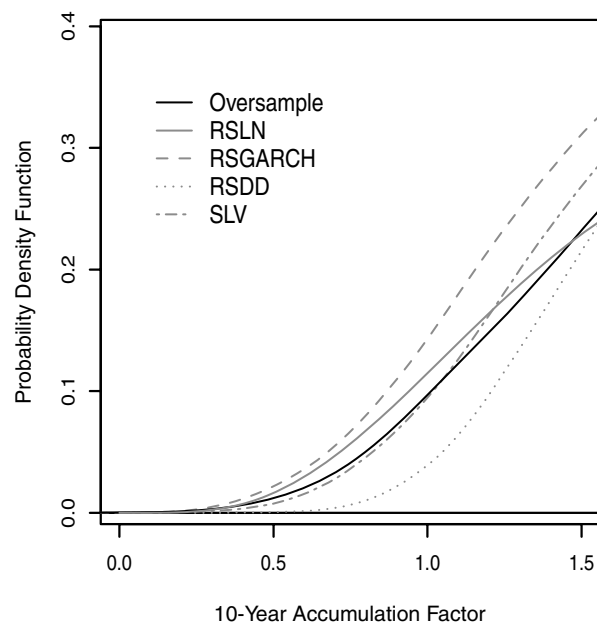
one-year. Once we move to the quantiles that are within the range of the data, oversampling gives broadly the same information as the regular bootstrap.

The left tail quantiles for the oversampled distribution and the model 10-year accumulation factors are given in Table 6. We generated 1,000 values for the 10-year accumulation factor, sampling from the original values in six-month blocks.

The table shows that only two models pass this test for left tail weight for all three quantiles: the RSLN and RSGARCH models. The relevant values are bolded in the table. The SLV model lies close to the oversample distribution. All the rest are significantly thinner tailed than the oversample distribution. The relative weights are illustrated in Figure 6, which shows the left tails of the density functions for the oversample distribution and four of the model pdfs.

This test gives strong support that the RSLN and RSGARCH results for capital requirements in Table 2 are more reliable than the RSDD, despite the apparent similarity in the distributions in Section 4.

Figure 6
Left Tail of Distribution of 10-Year Accumulation Factors



The same arguments that we have applied to the left tail also apply to the right tail. In Figure 7 we compare the same four distributions with the oversample distribution. We see that only one distribution is fatter tailed on the right side, the RSLN.

8. COMMENTS AND CONCLUSION

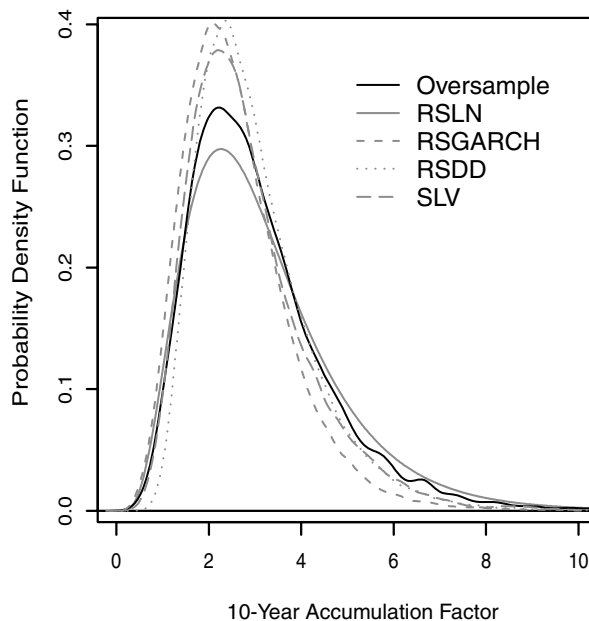
In this paper we have used the maximum data period we consider feasible without taking in the 1940s, a period of strange economic (and other) conditions. The parameter estimation for most of the models will be fairly sensitive to the period of data selected for estimation. This is particularly true for the regime-switching models, where the more volatile regime is rarely visited, meaning fairly rare historical observations. As mentioned in the introduction, if the data period is too short, practically any sensible model will fit equally well.

It should not be assumed that stable parameter estimation implies a better model fit. The lognormal parameters are fairly stable over different data periods, but the historical fit is bad over the same periods. However, the uncertainty inherent in the estimated parameters is an important consideration. More discussion of this issue is available in Hardy (2003), including an explanation of how to use a Bayesian predictive distribution to quantify the effect of parameter uncertainty on actuarial estimates.

While maximum likelihood arguments can assess whether one model has a better overall fit to the data than another, they do not tell us whether any of the models provide an adequate fit. We have shown that the residuals give some further information on this, and for the monthly S&P data set, the regime-switching models appear to provide a substantially better fit than the ARCH type or stochastic log volatility models. However, this does not necessarily help to determine capital requirements since we show that, using the actuarial approach, two very similar models, the RSLN and RSDD, can give very different capital requirements for a simple example variable annuity benefit.

Using an extension of the bootstrap has given us a new tool to explore the tail fit for the accumulation factors, by extracting more information from the data. Here we show again that the RSLN and RSGARCH models appear to offer a satisfactory fit to the left tail for the 10-year accumulation factors,

Figure 7
Full Distribution of 10-Year Accumulation Factors



while the RSDD model, along with the MARCH, ILN, and GARCH models, does not. The SLV is borderline on this test.

ACKNOWLEDGMENTS

This work was supported by the Natural Science and Engineering Research Council of Canada, by the Actuarial Foundation of Canada in conjunction with the Actuarial Education and Research Fund of the Actuarial Foundation, and by the Institute of Quantitative Finance and Insurance at the University of Waterloo.

REFERENCES

- AMERICAN ACADEMY OF ACTUARIES (AAA). 2005. Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Annuities and Similar Products, Presented by the American Academy of Actuaries Life Capital Adequacy Subcommittee to the National Association of Insurance Commissioners Capital Adequacy Task Force (C3 Phase 2 Report). www.actuary.org/pdf/life/c3_june05.pdf.
- BAKSHI, GURDIP, CHARLES CAO, AND ZHIWU CHEN. 1997. Empirical Performance of Alternative Option Pricing Models. *Journal of Finance* 52: 2003–49.
- . 2000. Pricing and Hedging Long-Term Options. *Journal of Econometrics*, 94: 277–318.
- BLACK, FISCHER, AND MYRON SCHOLES. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–54.
- CAMPBELL, JOHN Y. ANDREW W. LO, AND A. CRAIG MACKINLAY. 1996. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- CANADIAN INSTITUTE OF ACTUARIES (CIA). 2001. *Report of the Task Force on Segregated Fund Investment Guarantees*. Canadian Institute of Actuaries; available from www.actuaries.ca/publications/2002/202012e.pdf.
- DUAN, JUN CHUAN. 1995. The GARCH Option Pricing Model. *Mathematical Finance* 5: 13–32.
- EFRON, BRADLEY, AND ROBERT J. TIBSHIRANI. 1993. *An Introduction to the Bootstrap*. Boca Raton, FL: Chapman and Hall/CRC Press.
- ENGLE, ROBERT F., ed. 1995. *ARCH: Selected Readings*. Oxford: Oxford University Press.
- GRAY, STEPHEN F. 1996. Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process. *Journal of Financial Economics* 42: 27–62.
- HAMILTON, JAMES D. 1989. A New Approach to the Economic Analysis of Non-stationary Time Series. *Econometrica* 57: 357–84.
- HARDY, MARY R. 2001. A Regime Switching Model of Long-Term Stock Returns. *North American Actuarial Journal* 5(2): 41–53.
- . 2003. *Investment Guarantees; Modeling and Risk Management for Equity Linked Life Insurance*. New York: Wiley.
- JARQUE, CARLOS M., AND ANIL K. BERA. 1980. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters* 6: 255–59.
- . 1987. A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 55: 163–72.
- PANNETON, CHRISTIAN-MARC. 2002. Mean-Reversion in Equity Models in the Context of Actuarial Provisions for Segregated Fund Investment Guarantees. *Segregated Funds Symposium Proceedings*. Canadian Institute of Actuaries.
- SHEPARD, NEIL. 2005. *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- WILKIE, A. DAVID. 1995. More on a Stochastic Asset Model for Actuarial Use. *British Actuarial Journal* 1(5): 777–964.
- WONG, ALBERT C. S., AND WAI-SUM CHAN. 2005. Mixture Gaussian Time Series Modeling of Long-Term Market Returns. *North American Actuarial Journal* 9(4): 83–94.

Additional discussions on this paper can be submitted until April 1, 2007. The authors reserve the right to reply to any discussion. Please see the Submission Guidelines for Authors on the inside back cover for instructions on the submission of discussions.