North American Actuarial Journal

ISSN: 1092-0277 (Print) 2325-0453 (Online) Journal homepage: www.tandfonline.com/journals/uaaj20

A Regime-Switching Model of Long-Term Stock Returns

Mary R.HardyA.S.A., F.I.A.

To cite this article: Mary R.HardyA.S.A., F.I.A. (2001) A Regime-Switching Model of Long-Term Stock Returns, North American Actuarial Journal, 5:2, 41-53, DOI: 10.1080/10920277.2001.10595984

To link to this article: https://doi.org/10.1080/10920277.2001.10595984

Published online: 04 Jan 2013.

A Regime-Switching Model of Long-Term Stock Returns*

Mary R. Hardy†

Abstract

In this paper I first define the regime-switching lognormal model. Monthly data from the Standard and Poor's 500 and the Toronto Stock Exchange 300 indices are used to fit the model parameters, using maximum likelihood estimation. The fit of the regime-switching model to the data is compared with other common econometric models, including the generalized autoregressive conditionally heteroskedastic model. The distribution function of the regime-switching model is derived. Prices of European options using the regime-switching model are derived and implied volatilities explored. Finally, an example of the application of the model to maturity guarantees under equity-linked insurance is presented. Equations for quantile and conditional tail expectation (Tail-VaR) risk measures are derived, and a numerical example compares the regime-switching lognormal model results with those using the more traditional lognormal stock return model.

1. Introduction

Traditional models for stock returns, including the original Black-Scholes approach, assume that returns follow a geometric Brownian motion. This implies that over any discrete time interval the return on stocks is lognormally distributed and that returns in nonoverlapping intervals are independent; that is, if S_t is the stock price at time t, then

$$\log \frac{S_t}{S_r} \sim N(\mu(t-r), \sigma^2(t-r))$$

for some μ and volatility σ. This independent lognormal (ILN) model is simple and tractable and provides a reasonable approximation over shorter time intervals, but it is less appealing for longer-term problems. Empirical studies indicate in particular that this model fails to capture more extreme price movements and stochastic variability in the volatility parameter.

A simple way to incorporate stochastic volatility is to assume that volatility takes one of K discrete values, switching between these values randomly. This is the basis of the regime-switching lognormal process (RSLN). This approach maintains some of the attractive simplicity of the ILN model but more accurately captures the more extreme observed behavior. The subject of this paper is a Markov regime-switching lognormal model. Regime switching allows the stock price process to switch between K regimes randomly; each regime is characterized by different model parameters, and the process describing which regime the price process is in at any time is assumed to be Markov (that is, the probability of changing regime depends only on the current regime, not on the history of the process).

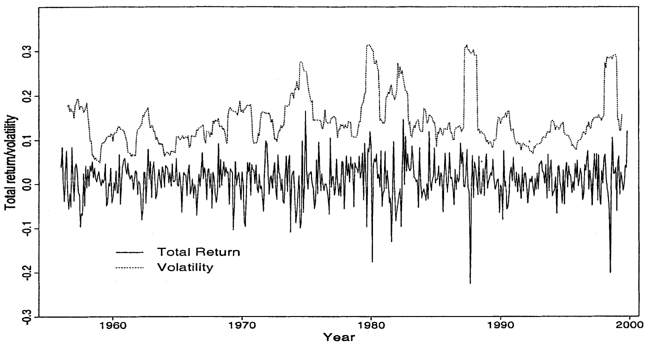
The rationale behind the regime-switching framework is that the market may switch from time to time between, say, a stable low-volatility state and a more unstable high-volatility regime. Periods of high volatility may arise, for example, because of short-term political or economic uncertainty.

Regime switching was introduced in Hamilton

*This research was supported by the National Science and Engineering Research Council of Canada. The TSE 300 data originate from the Statistics Canada CANSIM database. CANSIM is an official Mark of Statistics Canada. Parts of this work appeared in the paper "Stock Returns for Segregated Fund Models," commissioned by the Canadian Institute of Actuaries and presented to the Symposium on Stochastic Modeling for Segregated Funds/Variable Annuities Investment Guarantees.

†Mary Hardy, A.S.A., F.I.A., is an Associate Professor of Actuarial Science in the Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON Canada N2L 3G1, e-mail: mrhardy@uwaterloo.ca.

Figure 1
Monthly Total Returns and Annual Volatility, TSE 300



(1989), who described an autoregressive regime-switching process. In Hamilton and Susmel (1994) several regime-switching models are analyzed, varying the number of regimes and the form of the model within regimes. Their objective is to model various weekly econometric series; for these, the more complicated autoregressive conditionally heteroskedastic (ARCH)-type models within regimes seem to be necessary.

The simpler form that I consider in this paper, using lognormal distributions within regimes, appears to be sufficiently complex for the monthly total return data; details are given in Section 5.2. It is also mathematically tractable, as I show in Section 6. This model was also used by Bollen (1998), who constructed a lattice for valuing American options. He did not explore the empirical evidence for the model. Harris (1997) has developed a multivariate autoregressive regime-switching model for actuarial use, fitted to quarterly Australian data.

The objectives of this paper are the following:

- To explain briefly how to fit the model to the data, using a traditional likelihood approach (Section 4)
- To compare the fit of the RSLN model with other models in common use for both the Standard and Poor's (S&P) 500 and the TSE 300 total return indices (the TSE 300 index is the broad-based index of the Toronto Stock Exchange) (Section 5.2)
- To derive the distribution function for the RSLN model (Section 6)
- To derive the closed-form European option price formula using the RSLN model (Section 7)
- To show how the RSLN model may be applied to the calculation of risk measures for equity-linked insurance, and to compare the results with the more traditional lognormal model (Section 8).

2. The Data

Figure 1 shows monthly returns on the TSE 300 index, with dividends reinvested, together with estimated volatility, calculated using a 12-month moving standard deviation of the log returns. The data run from 1956 to 1999. The reason for the 1956 start date is that the TSE index was first introduced in January 1956. The S&P data cover the same period for ease of comparison. Figure 2 shows the same data for the S&P 500 index.

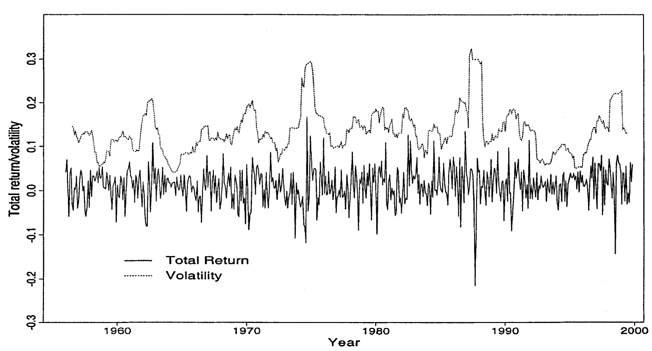
The best estimates (by maximum likelihood) of the annual volatility using these data sets are 15.63% for the TSE 300 data and 14.38% for the S&P 500 data. Mean monthly log returns are 0.8% for the TSE 300 and 0.9% for the S&P 500. The correlation coefficient between the two series is 0.773.

One observable feature of the data not captured by the lognormal model is volatility bunching, that is, periods of several months of high volatility, seen in both data sets in the middle 1970s and in the TSE data in the early 1980s. This feature is the one explicitly captured by the regime-switching approach.

3. The Model

Under the regime-switching lognormal model, it is assumed that the stock return process lies in

Figure 2
Monthly Total Returns and Annual Volatility, S&P 500



one of K regimes or states. Let ρ_t denote the regime applying in the interval $[t, t+1)$ (in months), $\rho_t = 1, 2, \ldots, K$, and S_t be the total return index value at t; then

$$\log \frac{S_{t+1}}{S_t}\bigg|\rho_t \sim N(\mu_{\rho_t}, \sigma^2_{\rho_t}).$$

I have investigated two- and three-regime models (that is, $K = 2, 3$) and have found no significant improvement in fit for the TSE data set from adding the third regime, and only a marginal improvement for the S&P data set; further details are given in Section 5.2. In most of this paper a two-regime model is used, that is, $K = 2$, which substantially simplifies the model and estimation compared with higher values for K. Hamilton and Susmel (1994), looking at weekly data (from 1962 to 1987) and assuming ARCH models for returns within each state, found some evidence for using three regimes, adding a very low volatility regime applying for a single period of the early 1960s; Harris (1997), using quarterly data and assuming autoregressive (AR) models within each regime, found no evidence for using more than two regimes.

The transition matrix $\boldsymbol{P}$ denotes the probabilities of moving regimes, that is,

$$p_{ij} = \Pr[\rho_{t+1} = j | \rho_t = i] \quad i = 1, 2, j = 1, 2.$$

Thus, for the two-regime (conditionally) independent lognormal model we have six parameters to estimate, $\Theta = \{\mu_1, \mu_2, \sigma_1, \sigma_2, p_{1,2}, p_{2,1}\}$.

4. MAXIMUM LIKELIHOOD ESTIMATION

4.1. Calculating the Likelihood Function

Let $Y_t = \log(S_{t+1}/S_t)$ be the log return in the $t+1^{\text{th}}$ month. The likelihood for observations $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is

$$L(\Theta) = f(y_1|\Theta) f(y_2|\Theta, y_1) f(y_3|\Theta, y_1, y_2) \cdots f(y_n|\Theta, y_1, \ldots, y_{n-1}),$$

where f is the pdf for y. Hence, the contribution to the log-likelihood of the t-th observation is

$$\log f(y_t|y_{t-1}, y_{t-2}, \ldots, y_1, \Theta).$$

We can calculate this recursively (following Hamilton and Susmel 1994, for example), by calculating for each t:

$$\begin{aligned} f(\rho_t, \rho_{t-1}, y_t | y_{t-1}, \ldots, y_1, \Theta) &= p(\rho_{t-1}|y_{t-1}, \ldots, y_1, \Theta) \\ &\quad \times p(\rho_t|\rho_{t-1}, \Theta) f(y_t|\rho_t, \Theta). \end{aligned} \tag{1}$$

On the right-hand side of this equation,

$p(\rho_t|\rho_{t-1}, \Theta)$ is the transition probability between the regimes

$f(y_t|\rho_t, \Theta) = \phi((y_t - \mu_{\rho_t})/\sigma_{\rho_t})$, where ϕ is the standard normal probability density function, and

The probability function $p(\rho_{t-1}|y_{t-1}, y_{t-2}, \ldots, y_1, \Theta)$ is found from the previous recursion; it is equal to

$$\frac{\begin{array}{l}(f(\rho_{t-1}, \rho_{t-2} = 1, y_{t-1}|y_{t-2}, y_{t-3}, \ldots, y_1, \Theta) \\ \quad + f(\rho_{t-1}, \rho_{t-2} = 2, y_{t-1}|y_{t-2}, y_{t-3}, \ldots, y_1, \Theta)\end{array}}{f(y_{t-1}|y_{t-2}, y_{t-3}, \ldots, y_1, \Theta)}.$$

We can then calculate $f(y_t|y_{t-1}, y_{t-2}, \ldots, y_1, \Theta)$ as the sum over the four possible values of Equation (1), that is, for $\rho_t = 1, 2$ and $\rho_{t-1} = 1, 2$.

To start the recursion, we need a value (given Θ) for $p(\rho_0)$, which we can find from the invariant distribution of the regime-switching Markov chain. The invariant distribution $\pi = (\pi_1, \pi_2)$ is the unconditional probability distribution for the process. Under the invariant distribution π, each transition returns the same distribution; that is, $\pi P = \pi$, giving $\pi_1 p_{1,1} + \pi_2 p_{2,1} = \pi_1$ and $\pi_1 p_{1,2} + \pi_2 p_{2,2} = \pi_2$. Clearly $p_{1,1} + p_{1,2} = 1.0$, so that $\pi_1 = p_{2,1}/(p_{1,2} + p_{2,1})$, and similarly $\pi_2 = 1 - \pi_1 = p_{1,2}/(p_{1,2} + p_{2,1})$.

Hence, we can start the recursion by calculating for a given parameter set Θ:

$$f(\rho_1 = 1, y_1|\Theta) = \pi_1 \phi\left(\frac{y_1 - \mu_1}{\sigma_1}\right),$$

$$f(\rho_1 = 2, y_1|\Theta) = \pi_2 \phi\left(\frac{y_1 - \mu_2}{\sigma_2}\right),$$

$$f(y_1|\Theta) = f(\rho_1 = 1, y_1|\Theta) + f(\rho_1 = 2, y_1|\Theta),$$

and we calculate for use in the next recursion the two values of

$$p(\rho_1|y_1, \Theta) = \frac{f(\rho_1, y_1|\Theta)}{f(y_1|\Theta)}.$$

Maximizing the likelihood function over the six parameters may be done with standard search methods.

4.2. Maximum Likelihood Estimation Results

The maximum likelihood parameters for the TSE 300 and the S&P 500 data are given in Table 1, with approximate standard errors in parentheses. The parameters are fairly similar, as we would expect. In both cases the high-volatility regime has a negative mean and an annual volatility of approximately 25%. The main difference is in the probability of moving from the high-volatility regime to the low-volatility regime, which is estimated at 21% for the TSE data and at 37.1% for the S&P data, but in both cases the estimated standard errors are high.

This indicates that the persistence of the high-volatility regime appears less for the S&P data. The probability that a single run in regime i lasts t months is $p_{ii}^{t-1}p_{ij}$, so that the average length of a single run in regime i is $1/p_{ij}$ months. This gives an average run in the high-volatility regimes of 4.8 months for the TSE data and 2.6 months for the S&P data. In both cases the estimated standard errors for this parameter are high. Periods of high volatility are in practice often associated with falling markets. This is corroborated by the negative mean log return parameters in the high-volatility regimes.

Note that asymptotic results for the maximum likelihood estimation should not be relied on for inference for this data set. As the data are serially correlated under the RSLN, we cannot treat them as 527 independent observations; more accurately it should be viewed as a single, multivariate observation. This means that the "standard errors" quoted in the table should be regarded with caution. Reliable information about the uncertainty associated with these estimates is not available. Where this uncertainty is important it may be preferable to use a Bayesian approach to parameter estimation. In Hardy (1999) parameters are estimated using the Metropolis-Hastings algorithm (a form of the Markov Chain Monte Carlo methodology). This provides a sample from the joint posterior distribution for the parameters, which gives reliable information on the joint parameter uncertainty. This work incidentally supports the fact that the uncertainty about the $p_{2,1}$ parameters are very high, especially for the S&P data.

5. COMPARISON WITH OTHER MODELS

5.1. Introduction

The principle of parsimony indicates that more complex models require significant improvement in fit to be worthwhile. More complex here means using more parameters.

For models with an equal number of parameters it is appropriate to choose the model with the higher log-likelihood. For models with different numbers of parameters, common selection criteria are the likelihood ratio test (LRT), the Akaike information criterion (AIC) (Akaike 1974), and the Schwartz Bayes Criterion (SBC) (Schwartz 1978).

In this section all these tests are applied to the following models. In the description below, Y_t is the log-return in the $t + 1^{\text{th}}$ month.

1. ILN: the independent lognormal model described in Section 1, where
$$Y_t = \mu + \sigma\epsilon_t, \quad \epsilon_t \text{ iid} \sim N(0, 1).$$
The independent lognormal model is in common use for valuing embedded options, for example, in equity-linked contracts.

2. AR(1): A first-order autoregressive model, where
$$Y_t = \mu + a(Y_{t-1} - \mu) + \sigma\epsilon_t, \quad \epsilon_t \text{ iid} \sim N(0, 1).$$

Table 1

Maximum Likelihood Parameters, with Estimated Standard Errors

TSE 300		
$\hat{\mu}_1 = 0.0123$ (0.002) $\hat{\mu}_2 = -0.0157$ (0.010)	$\hat{\sigma}_1 = 0.0347$ (0.001) $\hat{\sigma}_2 = 0.0778$ (0.009)	$\hat{p}_{1,2} = 0.0371$ (0.012) $\hat{p}_{2,1} = 0.2101$ (0.086)
S&P 500		
$\hat{\mu}_1 = 0.0126$ (0.002) $\hat{\mu}_2 = -0.0185$ (0.014)	$\hat{\sigma}_1 = 0.0350$ (0.001) $\hat{\sigma}_2 = 0.0748$ (0.009)	$\hat{p}_{1,2} = 0.0398$ (0.015) $\hat{p}_{2,1} = 0.3798$ (0.123)

The AR(1) model allows for serial correlation in the data.

3. ARCH(1): The autoregressive conditionally heteroskedastic model, where the variance is a function of the evolving process:

$$Y_t = \mu + \sigma_t \epsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 (Y_{t-1} - \mu)^2.$$

The autocorrelations in the ARCH model are all zero. We can combine the AR and ARCH models for the process:

$$Y_t = \mu + a(Y_{t-1} - \mu) + \sigma_t \epsilon_t, \quad \epsilon_t \text{ iid} \sim N(0, 1),$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 (Y_{t-1} - \mu)^2.$$

In Table 2 these are referred to as ARCH and AR-ARCH, respectively.

4. GARCH(1, 1): The generalized autoregressive conditionally heteroskedastic model:

$$Y_t = \mu + \sigma_t \epsilon_t,$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 (Y_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2.$$

The GARCH process, like the ARCH, has zero autocorrelation. Again, we can combine the AR(1) model with the GARCH for the process:

$$Y_t = \mu + a(Y_{t-1} - \mu) + \sigma_t \epsilon_t, \quad \epsilon_t \text{ iid} \sim N(0, 1),$$
$$\sigma_t^2 = \alpha_0 + \alpha_1 (Y_{t-1} - \mu)^2 + \beta \sigma_{t-1}^2.$$

In Table 2 these two models are denoted GARCH and AR-GARCH, respectively.

5. RSAR(1) is a two-regime version of the AR(1) model, that is,

$$Y_t | \rho_t = \mu_{\rho_t} + a_{\rho_t}(Y_{t-1} - \mu_{\rho_t}) + \sigma_{\rho_t} \epsilon_t, \quad \epsilon_t \text{ iid} \sim N(0, 1), \rho_t = 1, 2.$$

The RSLN model introduces autocorrelation through the regime process. The RSAR model should capture remaining autocorrelation.

6. RSLN-3 is a three-regime lognormal model.

Both the ARCH and GARCH models allow for the volatility of the process to vary and are designed to model periods of high and low volatility in financial series. ARCH models were introduced in Engle (1982), and Bollerslev (1986) extended these to the GARCH formulation. A comprehensive text on these models is Hamilton (1994).

5.2. Selection Criteria

5.2.1. The Likelihood Ratio Test

The likelihood ratio test (see, for example, Klugman, Panjer, and Willmot 1998) compares embedded models, that is, where a model with k_1 parameters is a special case of a more complex model with $k_2 > k_1$ parameters. Let l_1 be the log-likelihood of the simpler model, and l_2 be the log-likelihood of the more complex model. The test statistic is $2(l_2 - l_1)$. The null hypothesis is

H_0 : No significant improvement in Model 2.

Under the null hypothesis, the test statistic has χ^2 distribution, with degrees of freedom equal to the difference between the number of parameters in the two models.

Not all of the models we consider are embedded; if we denote embeddedness by $\subset$, we have ILN $\subset$ RSLN-2 $\subset$ RSLN-3 and RSLN-2 $\subset$ RSAR(1). However, even where models are not embedded, the likelihood ratio test can be used for model selection, although the χ^2 distribution is in this case only an approximation.

In Table 2 the final column gives the p-value for a likelihood ratio test of the RSLN model against each of the other models listed. For models with fewer than six parameters, the null hypothesis is that the simpler model is a "better" fit than the RSLN. Low p-values indicate rejection of the null hypothesis. Comparing the two-regime RSLN-2 model with models with more than six parameters, acceptance of the null hypothesis (high p-value) implies acceptance of the RSLN-2 model.

5.2.2. The Akaike Information Criterion

The Akaike Information Criterion (AIC) uses the model that maximizes $l_j - k_j$, where l_j is the log-likelihood under the jth model and k_j is the number of parameters. Using this criterion, each extra parameter must improve the log-likelihood by at least one. This criterion was derived heuristically by Akaike (1974). It is popular for ease of application but is not rigorously founded. It captures in the simplest possible way the intuition that, from principle of parsimony, each extra parameter added must be worthwhile in terms of the log-likelihood improvement.

5.2.3. The Schwartz Bayes Criterion

The Schwartz Bayes Criterion uses the model that maximizes $l_j - \frac{1}{2} k_j \log n$, where n is the

Table 2
Comparison of Selection Information for Lognormal, Autoregressive, and Regime Switching Models

Model (j)	Parameters (k_j)	log L (l_j)	SBC ($l_j - \frac{1}{2} k_j \log n$)	AIC ($l_j - k_j$)	LRT (p)
TSE 300 (1956–99 Monthly Total Returns)					
ILN	2	885.6	879.4	883.6	$<10^{-8}$
AR(1)	3	887.4	878.0	884.4	$<10^{-8}$
ARCH	3	888.5	879.1	885.5	$<10^{-8}$
AR-ARCH	4	890.8	878.3	887.8	$<10^{-8}$
GARCH	4	896.0	883.5	892.0	$<10^{-8}$
AR-GARCH	5	897.9	882.2	894.9	$<10^{-8}$
RSLN-2	6	922.7	**903.9**	**917.7**	
RSAR-2	8	922.9	897.8	914.9	0.82
RSLN-3	12	925.9	888.3	913.9	0.38
S&P 500 (1956–99 Monthly Total Returns)					
ILN	2	929.4	923.1	927.4	$<10^{-8}$
AR(1)	3	929.6	917.1	926.6	$<10^{-8}$
ARCH	3	933.3	923.9	930.3	$<10^{-8}$
AR-ARCH	4	933.4	920.9	929.4	$<10^{-8}$
GARCH	4	938.9	926.4	934.9	$<10^{-6}$
AR-GARCH	5	939.1	923.4	934.1	$<10^{-6}$
RSLN-2	6	952.5	**933.6**	946.5	
RSAR-2	8	952.6	927.5	944.6	0.91
RSLN-3	12	960.8	923.2	**948.8**	0.01

sample size. For a sample of 527 (corresponding to the monthly data 1956–99) each additional parameter must increase the log-likelihood by at least 3.1. This criterion was derived by Schwartz (1978). Like the AIC, new parameters must be worthwhile in terms of likelihood improvement, but in this case the improvement depends on the amount of data available: extra parameters are penalized more heavily where the sample size is large.

5.3. Results, TSE and S&P Data

Table 2 shows that the RSLN-2 model provides a significant improvement over all other models for the TSE data using each of the three selection criteria. For the S&P data, selection is not quite so definite. According to the likelihood ratio test and the AIC, there is a marginal improvement in fit from using three regimes. The third regime is an ultra-low volatility regime with transitions to and from the low-volatility regime only. The Schwartz Bayes criterion still favors the two-regime model. Given the added complexity of the three-regime model, and the marginal nature of the improvement, I pursue the two-regime version of the model in this paper. However, all the topics discussed subsequently can be adapted for the three-regime version of the model.

A longer S&P data series (price index) has also been fitted to the same two- and three-regime models. The data run from 1926 to 1998. Under the two-regime model, the maximum likelihood parameters for the long data set are similar for the first regime, but the variance and persistence of the second regime are much higher than the parameters found using postwar data. The volatility for the second regime is estimated at 12% per month, and the probability of moving from the high-volatility regime to the low-volatility regime is estimated at 0.1. This effect arises from the prolonged period of very high volatility in the 1930s. Once again, there is a marginal improvement in fit from using three regimes (p-value for the likelihood ratio test is 0.02).

5.4. October 1987

A common concern where a model for stock returns is proposed is whether the model captures the sort of extreme value observed in October 1987, when the TSE 300 log-return was -0.2552. Using the post-1956 parameters, under a monthly lognormal model, this is six standard deviations away from the mean. The expected number of observations this small appearing in a sample of 527 observations is approximately 2×10^{-6}. In other words, an observation this small would appear in only one in approximately 700,000 samples of 527 values.

Under the RSLN model, given that the process is in the high-volatility regime (regime-2), the observation is 3.078 standard deviations away from the mean; the probability of an observation at least as small as this within regime-2 is 0.104%. Allowing for the probability of being in regime-2 reduces the probability for the individual observation to 0.016%. The probability of such an observation in a sample of 527 is approximately 8%: that is, under the RSLN model, around 1 in 12 samples of 527 monthly observations would include a value at least as small as the October 1987 value. Although the October 1987 value is a rare observation under the RSLN model, with greater than 5% probability of one such value in a sample, it is not nearly sufficiently extreme to reject the model.

Note that monthly data and monthly models are used; we cannot infer probabilities associated with weekly or daily stock movements from the monthly RSLN model.

6. USING THE TWO-REGIME RSLN MODEL

6.1. Probability Function for Total Sojourn in Regime-1

In applying the regime-switching model it is very useful to have a probability distribution for the total number of months spent in regime-1. We can use this probability function to calculate the distribution function, density function, or moments of the stock price process S_n.

Let R be the total number of months spent in regime-1; $R \in \{0, 1, \ldots, n\}$. Denote the probability function $\Pr[R = r]$ by $p(r)$. Let R_t be the total sojourn in regime-1 in the interval $[t, n)$, and consider $\Pr[R_t = r|\rho_{t-1}]$, for $r = 0, 1, \ldots, n - t$ and $t = 1, \ldots, n - 1$. Clearly $\Pr[R_t = r|\rho_{t-1}] = 0$ for $r > n - t$ or $r < 0$.

For example, $\Pr[R_{n-1} = 0|\rho_{t-1} = 1]$ is the probability that the last time unit is not spent in regime-1 given that the process is in regime-1 in the previous period—that is, for $t \in [n - 2, n - 1)$—so that $\Pr[R_{n-1} = 0|\rho_{t-1} = 1] = p_{1,2}$. Similarly,

$$\Pr[R_{n-1} = 1|\rho_{t-1} = 1] = p_{1,1},$$

$$\Pr[R_{n-1} = 0|\rho_{t-1} = 2] = p_{2,2},$$

$$\Pr[R_{n-1} = 1|\rho_{t-1} = 2] = p_{2,1}.$$

We can work backwards from these values to the required probabilities for $R = R_0$ using the relationship

$$\begin{aligned}\Pr[R_t = r|\rho_{t-1} = 1] &= p_{\rho_{t-1},1} \Pr[R_{t+1} = r - 1|\rho_t = 1] \\ &\quad + p_{\rho_{t-1},2} \Pr[R_{t+1} = r|\rho_t = 2]. \end{aligned} \tag{2}$$

The justification for this is that, immediately after the transition at time t, either the process is in regime-1 (i.e., $\rho_t = 1$, with probability $p_{\rho_{t-1},1}$), which leaves $r - 1$ time periods to be spent in regime-1 subsequently, or the process is in regime-2 (i.e., $\rho_t = 2$ with probability $p_{\rho_{t-1},2}$), in which case r time periods must be spent in regime-1 in the interval $[t + 1, n)$.

Ultimately we can find the probability functions for R_0 conditional on regime-1 as the starting point, $\Pr[R_0 = r|\rho_{-1} = 1]$, and conditional on regime-2 as a starting point, $\Pr[R_0 = r|\rho_{-1} = 2]$. Using the stationary distribution for the regimes, we can then find the probability function of R_0 as

$$\begin{aligned}\Pr[R_0 = r] = p(r) &= \pi_1 \Pr[R_0 = r|\rho_{-1} = 1] \\ &\quad + \pi_2 \Pr[R_0 = r|\rho_{-1} = 2]. \end{aligned} \tag{3}$$

6.2. Probability Functions for S_n

Using the probability function for R, the distribution of the total return index at time n can be calculated analytically. Let S_n represent the total return index at n; assume $S_0 = 1$. Then

$$S_n|R \sim \text{lognormal}(\mu^*(R), \sigma^*(R)),$$

$$\text{where } \mu^*(R) = R\mu_1 + (n - R)\mu_2, \tag{4}$$

$$\sigma^*(R) = \sqrt{R\sigma_1^2 + (n - R)\sigma_2^2}. \tag{5}$$

Then, if $p(r)$ is the probability function for R,

$$F_{S_n}(x) = \Pr[S_n \leq x] = \sum_{r=0}^{n} \Pr[S_n \leq x|R = r]p(r) \tag{6}$$

$$= \sum_{r=0}^{n} \Phi\left(\frac{\log x - \mu^*(r)}{\sigma^*(r)}\right) p(r), \qquad (7)$$

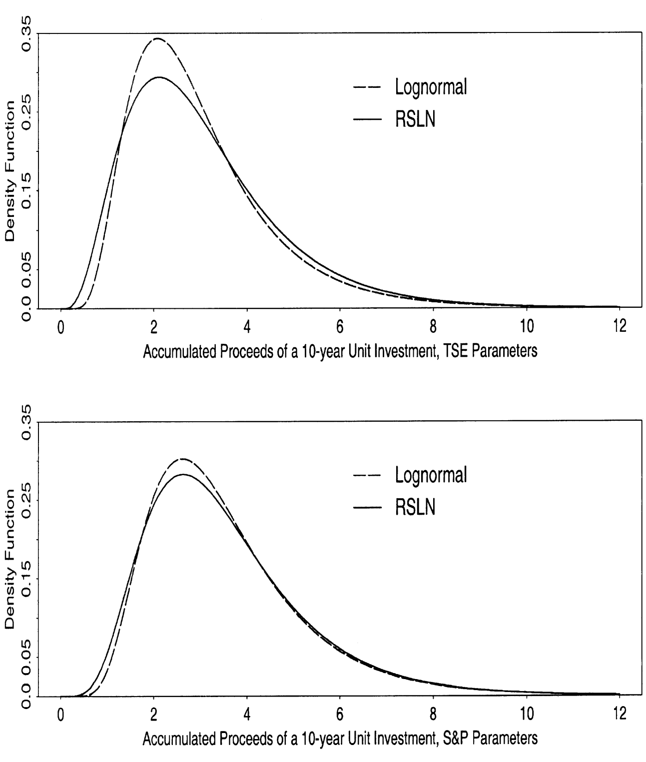
where $\Phi(\ )$ is the standard normal probability distribution function. Similarly, the probability density function for S_n is

$$f_{S_n}(x) = \sum_{r=0}^{n} \phi\left(\frac{\log x - \mu^*(r)}{\sigma^*(r)}\right) p(r), \qquad (8)$$

where $\phi(\ )$ is the standard normal density function.

Equation (8) has been used to calculate the density functions shown in Figure 3, which shows the RSLN and lognormal density functions for the stock price at $t = 10$ years, given $S_0 = 1.0$, using both the TSE and S&P parameters. In both cases, over this long term, the left tail is substantially fatter for the RSLN model than for the lognormal model. This has important implications for longer-term actuarial applications. For example, in modeling the maturity guarantees in Canadian segregated fund contracts, the lognormal model has been very popular (see, for example, the report of the Task Force on Segregated Funds (TFSF) 2000). The stock price model is, typically, being used with a 20-year horizon for modeling the risks for these contracts. For these long terms the fatter tail of the RSLN model will have a substantial effect on the results. This is discussed further in Section 8.

Figure 3
Probability Density Curves for ILN and RSLN Models, TSE and S&P Data



The probability function for the sojourn times can also be used to find unconditional moments of the stock price at any time n:

$$\begin{aligned}
E[(S_{t+1})^k] &= E[E[(S_{t+1})^k | R]] \\
&= E\left[\exp(k(R\mu_1 + (n-R)\mu_2) + \frac{k^2}{2}(R\sigma_1^2 + (n-R)\sigma_2^2))\right] \\
&= E\left[\exp\left(R\left(k(\mu_1 - \mu_2) + \frac{k^2}{2}(\sigma_1^2 - \sigma_2^2)\right)\right)\right] \times \exp\left(kn\mu_2 + \frac{k^2}{2} n\sigma_2^2\right) \\
&= \exp\left(kn\mu_2 + \frac{k^2}{2} n\sigma_2^2\right) \times \sum_{r=0}^{n} \exp\left(r\left(k(\mu_1 - \mu_2) + \frac{k^2}{2}(\sigma_1^2 - \sigma_2^2)\right)\right) p(r).
\end{aligned}$$

7. OPTION PRICING

In the conventional Black-Scholes framework, where the asset price S_n has a lognormal distribution, the Black-Scholes price for a put option with strike K, maturing at n, valued at time $t = 0$, is

$$\begin{aligned}
BSP &= e^{-rn} E_Q[\max(K - S_n, 0)] \\
&= Ke^{-rn} N(-d_2) - S_0 N(-d_1),
\end{aligned}$$

where r is the risk-free rate (force) of interest, N denotes the standard normal distribution function, S_t is the asset price at t, and

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)n}{\sigma\sqrt{n}}, \quad d_2 = d_1 - \sigma\sqrt{n}.$$

The parameter σ is the volatility of the asset return, which corresponds to the lognormal parameter from the ILN model. The subscript Q indicates that the expectation is with respect to the risk-neutral measure Q.

Bollen (1998) uses the regime-switching model to price American and European options: American by a lattice method, European by simulation. Since the market is incomplete in a regime-switching model, the resulting Q measure is not uniquely determined. Bollen uses a Q-measure under which the transition probabilities are unchanged, the move from P-measure to Q-measure being effected by changing the log mean parameter in regime-1 from μ_1 to $r - \sigma_1^2/2$ and in regime-2 from μ_2 to $r - \sigma_2^2/2$, where r is the risk-free force of interest.

In fact, it is not necessary to use simulation. We can calculate the European option price directly, using the probability distribution for R. Conditional on knowing R, the asset price $S_n|R$ has a lognormal distribution, with parameters dependent on R. Thus, the put option price under this Q-measure is

$$\mathrm{E}_R[e^{-rn}\mathrm{E}_Q[\max(K - S_n, 0)|R]] = \mathrm{E}_R[BSP(R)],$$

where

$$BSP(R) = Ke^{-rn}N(-d_2(R)) - S_0N(-d_1(R)),$$

$$d_1(R) = \frac{\ln(S_0/K) + (nr + (R(\sigma_1^2/2) + (n - R)(\sigma_2^2/2)))}{\sqrt{(R\sigma_1^2 + (n - R)\sigma_2^2)}},$$

$$d_2(R) = d_1(R) - \sqrt{(R\sigma_1^2 + (n - R)\sigma_2^2)}.$$

Table 3 shows some option prices calculated for various strike prices, using the TSE parameters. Figures are given for a one-year put option and a 10-year put option. Such longer options are common in insurance applications, in particular for guarantees under segregated fund contracts. Figures for the implied Black-Scholes volatility are given in the final column. The assumed risk free rate is 6% pa. All figures are per 100 initial asset price. In Table 4 the same figures are shown using the S&P parameters.

The implied volatilities are shown graphically in Figure 4. The strike prices quoted are for an initial stock price of 100. It can be seen that the regime-switching model gives a lopsided volatility smile (or smirk); that is, the implied volatility for the at-the-money option is smaller than the implied volatility for out-of-the-money and in-the-

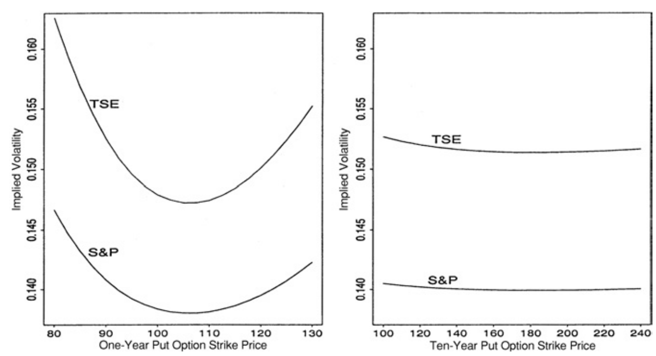
Table 3
Put Option Prices and Black-Scholes Implied Volatility, TSE Parameters

Strike Price K	Option Price	Implied Volatility
One-Year Put Option Prices		
80	0.232	16.25%
100	3.275	14.79
120	14.876	15.01
Ten-Year Put Option Prices		
100	1.800	15.27
180	18.198	15.14
260	50.212	15.18

Table 4
Put Option Prices and Black-Scholes Implied Volatility, S&P Parameters

Strike Price K	Option Price	Implied Volatility
One-Year Put Option Prices		
80	0.130	14.67%
100	2.938	13.84
120	14.563	13.95
Ten-Year Put Option Prices		
100	1.322	14.05
180	16.803	13.99
260	48.938	14.02

Figure 4
Black-Scholes Implied Volatility Smirk



money options, both for one-year and for 10-year options, though the effect is much more marked for the shorter term. ("At-the-money" for a 10-year option is equivalent to a strike of $K = 100e^{10r} = 182$ allowing for discounting.) This is a phenomenon often observed in practice.

Although the difference between the S&P and TSE figures appear substantial, given the large approximate standard error for the S&P transition probability out of the high-volatility regime, the difference may not be significant.

8. Risk Measures for Segregated Fund Contracts

8.1. Segregated Fund Contracts

The problem of modeling the maturity guarantee liability under equity-linked or segregated fund contracts was a major driving force behind this exploration of the RSLN model. Segregated fund contracts are a form of equity-linked insurance that has proved very popular with insurers and consumers in Canada in the last few years. The Task Force on Segregated Funds, established by the Canadian Institute of Actuaries (CIA) to investigate provisions for the liabilities arising, found that the lognormal distribution was a popular assumption for the underlying stock returns (TFSF 2000). Another was the Wilkie investment model (Wilkie 1995). It can be demonstrated (see, for example, Hardy 1999), that the total stock price returns modeled using the Wilkie model are very similar to a lognormal model with parameters fitted from the same data. In this section, risk measures for a simple segregated fund liability are calculated, comparing the results using a lognormal distribution of stock returns with those found using the regime-switching lognormal distribution.

Under the simplest form of the segregated fund contract, a premium is invested in a mutual fund for, say, 10 years. Expenses are deducted monthly. At maturity the policyholder receives the proceeds of the mutual fund, with a guarantee that the payment will not be less than the original premium. The insurer's liability at maturity is then $\max[G - F, 0]$, where G is the guaranteed maturity value (typically 75% or 100% of the premium) and F is the fund at maturity. The benefit can be viewed as a European put option.

For these contracts any analysis of the potential costs requires a long-term model for stock returns that has a realistic fit in the left tail; that is, we need a reasonable model of the worst possible outcomes for the stock returns.

8.2. Quantile Risk Measure

One of the measures used to assess the risk is to look at percentiles of the liability distribution; this is essentially a "value-at-risk" (VaR) approach. This section compares quantiles of the liability under a segregated fund contract using a lognormal distribution for stock returns with those found using the regime-switching lognormal model.

For simplicity, ignore withdrawals in these calculations. Let $G = 100$ be the amount of the guarantee, and assume that expenses of m% per month continuously compounded are deducted from the fund. Let S_n denote the underlying asset value at time n in months. Assume $S_0 = 100$. Then, for a n-month contract, $F = S_n e^{-nm}$ and the liability at maturity is

$$X = \max(G - S_n e^{-nm}, 0).$$

Let $\xi = \Pr[S_n e^{-nm} > G]$. Then, for all $\alpha \le \xi$, the 100α% quantile of the liability distribution is $V_\alpha = 0$.

For $\alpha > \xi$, V_α is found from

$$F_{S_n}((G - V_\alpha)e^{nm}) = \alpha \tag{9}$$

$$\Rightarrow V_\alpha = G - e^{-nm} F_{S_n}^{-1}(1 - \alpha), \tag{10}$$

where F_{S_n} is the distribution function of the underlying stock value at time n. In the lognormal case we can use the inverse standard normal distribution function, $z_\alpha = \Phi^{-1}(\alpha)$ to give

$$V_\alpha = G - S_0 \exp\{-z_\alpha \sqrt{n}\sigma + n\mu - nm\}, \tag{11}$$

where μ, σ are the parameters of the lognormal distribution of monthly returns. For the RSLN model, use Equation (7) for the distribution function $F_{S_n}(\,)$.

8.3. Conditional Tail Expectation Risk Measure

The quantile or VaR risk measure has many problems in application. These are summarized in, for example, Artzner et al. (1999) and Wirch and Hardy (1999). The solution of Artzner et al. to the problem of finding a coherent risk measure is to use the conditional tail expectation (CTE), defined as the expected value of the loss given that the loss falls in the upper $(1 - \alpha)$ tail of the distribution. This gives better results than does the quantile measure when comparing risks, because the CTE utilizes the whole tail of the distribution beyond the quantile, rather than the single quantile point.

For a continuous loss distribution (or, more strictly, if $V_{\alpha+\epsilon} > V_\alpha$ for any $\epsilon > 0$), then the CTE with parameter α, $0 \le \alpha < 1$ is

$$CTE(\alpha) = \mathrm{E}[X|X > V_\alpha], \tag{12}$$

where V_α is defined as in Equation (10).

Note that this definition, though intuitively appealing, does not give suitable results where V_α falls in a probability mass (this will happen for the segregated fund example for $\alpha < \xi$, in which case $V_\alpha = 0$). In this case, the CTE with parameter α is calculated as follows. Find $\beta' = \max\{\beta : V_\alpha = V_\beta\}$, then

$$CTE(\alpha) = \frac{(1 - \beta')\mathrm{E}[X|X > V_\alpha] + (\beta' - \alpha)V_\alpha}{1 - \alpha}. \tag{13}$$

In either case, where the CTE is calculated by simulation, it is found by taking the mean of the worst $100(1 - \alpha)\%$ of the simulations.

The CTE has been proposed by the CIA Task Force on Segregated Funds as the required risk measure for determining total balance sheet provision in respect of segregated fund guarantee liabilities (TFSF 2000).

To derive the CTE formulas, assume first that $\alpha \ge \xi$, then

$$CTE(\alpha)$$

$$= \mathrm{E}[X|X > V_\alpha] \tag{14}$$

$$= \mathrm{E}[G - S_n e^{-nm}|S_n < (G - V_\alpha)e^{nm}] \tag{15}$$

$$= \frac{1}{1-\alpha}\int_0^{(G-V_\alpha)e^{nm}} (G - ye^{-nm}) f_{S_n}(y)\, dy \tag{16}$$

$$= \frac{1}{1-\alpha}\Big(GF_{S_n}((G - V_\alpha)e^{nm})$$

$$- \int_0^{(G-V_\alpha)e^{nm}} ye^{-nm} f_{S_n}(y)\, dy\Big) \tag{17}$$

$$= G - \frac{e^{-nm}}{1-\alpha}\int_0^{(G-V_\alpha)e^{nm}} y f_{S_n}(y)\, dy. \tag{18}$$

If $S_n \sim LN(n\mu, \sqrt{n}\sigma)$, then for $\alpha \ge \xi$,

$$CTE(\alpha)$$

$$= G - \frac{\exp(n\mu - nm + n\sigma^2/2)}{1-\alpha} \times \Phi\left(\frac{\log(G - V_\alpha) - n\mu - nm - n\sigma^2}{\sqrt{n}\sigma}\right) \tag{19}$$

$$= G - \frac{\exp(n\mu - nm + n\sigma^2/2)}{1-\alpha} \times \Phi(-z_\alpha - \sqrt{n}\sigma). \tag{20}$$

If $S_n \sim$ RSLN, then $S_n|R \sim$ lognormal$(\mu^*(R), \sigma^*(R))$, where $\mu^*(\ )$ and $\sigma^*(\ )$ are defined in Equations (4) and (5). It is straightforward to show that the CTE for the RSLN distribution is, for $\alpha \ge \xi$,

$$CTE(\alpha) = G - \frac{e^{-nm}}{1-\alpha}\sum_{r=0}^{n} p(r)\Bigg(e^{\mu^*(r)+\sigma^*(r)^2/2} \Phi\left(\frac{\log(G - V_\alpha) - \mu^*(r) - nm - \sigma^*(r)^2}{\sigma^*(r)}\right)\Bigg). \tag{21}$$

In general, for $\alpha < \xi$ use Equation (13), with $\beta' = \xi$ and $V_\alpha = V_\xi = 0$ so that $CTE(\xi) = E[X|X > 0]$ and

$$CTE(\alpha) = \frac{(1 - \xi)}{(1 - \alpha)} CTE(\xi). \tag{22}$$

So Equations (20) and (21) can be adapted for $\alpha < \xi$ using Equation (22).

Calculation of the CTE using either Equation (20) or (21) is relatively straightforward. Examples are given in the following section.

8.4. Numerical Comparison of Risk Measures for LN and RSLN Distributions

In Table 5, figures are given for quantile and CTE risk measures for a segregated fund contract that

Table 5
Risk Measure Values for 10-Year Segregated Fund Contract, TSE Parameters

Quantile Risk Measure V_α (Percentage of Fund Value)				
Model	ξ	$V_{0.90}$	$V_{0.95}$	$V_{0.975}$
Lognormal	0.9145	0.0	12.744	25.327
RSLN	0.8827	5.842	25.918	40.438

CTE Risk Measure $CTE(\alpha)$ (Percentage of Fund Value)			
Model	$CTE(0.90)$	$CTE(0.95)$	$CTE(0.975)$
Lognormal	16.117	27.918	37.229
RSLN	29.305	43.043	53.517

matures in 10 years; the guarantee is equal to the fund market value at the start of the projection. Management fees of 0.25% per month, compounded continuously, are deducted. Lapses and deaths are ignored for simplicity.

The effect of the increased probability in the left tail of the return distribution in Figure 3 comes through in the markedly higher figures for the RSLN model in Table 5. The ξ parameter gives the probability that the guarantee will have a non-zero cost. Under the lognormal model this is 91.45%; under the RSLN assumption it is slightly lower at 88.27%. More significant is the difference between the risk measures. If the actuary decides to hold capital at the 95% level of the loss distribution, the lognormal assumption requires 12.7% of the fund, whereas the RSLN assumption requires more than double this, at 25.9%. Recall that the RSLN model fits the data very significantly better than the lognormal model; it seems reasonable to infer that the lognormal distribution will understate the true risk for losses dependent on the left tail of the return distribution.

9. CONCLUSION

The regime-switching lognormal model captures observed stock return behavior, in particular extreme observations such as October 1987, and volatility bunching. Statistically, the RSLN model provides a significantly better fit to the TSE and S&P data than do other popular models, and, at least for the TSE data, the two-regime version of the model appears to be sufficient.

The RSLN model is mathematically tractable; using iterated expectation over the regime-1 sojourn time distribution, it is possible to calculate the distribution and density functions accurately. It is also simple to calculate Black-Scholes option prices using a "natural" Q-measure. The implied volatility curves resulting exhibit a volatility smile as seen in practice.

It is also possible to use the RSLN model to calculate quantiles and CTE measures for the losses under a segregated fund contract. This is an application where the lognormal model is a popular choice. Comparing the risk measures under the lognormal and RSLN models indicates that the lognormal model (with model parameters fitted from the data) produces substantially smaller values for quantile and CTE measures. Given the markedly better fit to the data of the RSLN model, it is possible that the risk measures calculated using the lognormal model substantially underestimate the true risk measures.

ACKNOWLEDGMENTS

Discussions on different areas related to this paper with Howard Waters and especially with Phelim Boyle were extremely helpful. I also acknowledge gratefully earlier discussions with Glen Harris on regime-switching models.

REFERENCES

AKAIKE, H. 1974. "A New Look at Statistical Model Identification." *IEEE Trans Aut Control* 19: 716–23.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH. 1999. "Coherent Measures of Risk." *Mathematical Finance* 9: 203–28.

BOLLEN, N. P. B. 1998. "Valuing Options in Regime Switching Models." *Journal of Derivatives* 6: 38–49.

BOLLERSLEV, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307–27.

ENGLE, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation." *Econometrica* 50: 987–1007.

HAMILTON, J. D. 1989. "A New Approach to the Economic Analysis of Non-stationary Time Series." *Econometrica* 57: 357–84.

HAMILTON, J. D. 1994. *Time Series Analysis.* Princeton: Princeton University Press.

HAMILTON, J. D. AND R. SUSMEL. 1994. "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics* 64: 307–33.

HARDY, M. R. 1999. "Stock Return Models for Segregated Fund Guarantees." Institute for Insurance and Pensions Research, University of Waterloo, Research Report 99-14. Presented to the Symposium on Stochastic Modeling for Variable Annuity/Segregated Fund Investment Guarantees, Toronto, September 1999.

HARRIS, G. R. 1997. "Regime Switching Vector Autoregressions: A Bayesian Markov Chain Monte Carlo Approach." *Proceedings of the 7th International AFIR Colloquium* 1: 421–50.

KLUGMAN, S., H. PANJER, AND G. WILLMOT. 1998. *Ross Models: From Data to Decisions.* New York: Wiley.

SCHWARTZ, G. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461–64.

TASK FORCE ON SEGREGATED FUNDS. 2000. *Report of the Task Force on Segregated Fund Investment Guarantees.* Ottawa: Canadian Institute of Actuaries.

WILKIE, A. D. 1995. "More on a Stochastic Asset Model for Actuarial Use." *British Actuarial Journal* 1: 777–964.

WIRCH, J. L., AND M. R. HARDY. 1999. "A Synthesis of Risk Measures." *Insurance: Mathematics and Economics* 25: 337–48.

Discussions on this paper can be submitted until October 1, 2001. The author reserves 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.