Seeing Beyond Risk

Canadian Institute of Actuaries | Institut canadien des actuaires

Voir au-delà du risque

Research Paper

Calibration of Equity Returns for Segregated Fund Liabilities

Committee on Life Insurance Financial Reporting

February 2012

Document 212004

Ce document est disponible en français

Research papers do not necessarily represent the views of the Canadian Institute of Actuaries. Members should be familiar with research papers. Research papers do not constitute Standards of Practice and therefore are not binding. Research papers may or may not be in compliance with Standards of Practice. Responsibility for the manner of application of Standards of Practice in specific circumstances remains that of the members in the life insurance practice area.

Seeing Beyond Risk
Canadian Institute of Actuaries
Institut canadien des actuaires
Voir au-delà du risque

Memorandum

To: All Fellows, Affiliates, Associates and Correspondents of the Canadian Institute of Actuaries

From: Phil Rivard, Chair
Practice Council

Edward Gibson
Committee on Life Insurance Financial Reporting

Alexis Gerbeau, Chair
Designated Group

Date: February 3, 2012

Subject: **Research Paper: Calibration of Equity Returns for Segregated Fund Liabilities**

The Committee on Life Insurance Financial Reporting has created a designated group to draft this paper to provide support for an updated promulgation of calibration criteria for investment returns with respect to the valuation of segregated fund guarantees.

The existing calibration criteria for equity returns were developed using the data from January 1956 to December 1999. The data have been updated to cover the period from January 1956 to June 2010 for developing the updated calibration criteria.

The existing calibration criteria provided the maximum values for the accumulation factors for the 2.5^{th}, 5^{th}, and 10^{th} percentiles for the one-year, five-year, and 10-year horizons. A range for the mean of the one-year accumulation factor was also provided, as well as a minimum value for the standard deviation of the one-year accumulation factor. Updated figures for all these calibration criteria are presented in this paper. Maximum values for the 20-year accumulation factor have been added because of the growing popularity of longer-term products.

In addition, minimum values for the differences between the 90^{th}, 95^{th}, and 97.5^{th} percentiles and the median of the one-year accumulation factor have been established to capture the increasing risk caused by ratchet/reset and lock-in features that have become more common in products now being sold.

Finally, calibration criteria now cover more than broad-based Canadian equity indices. Two sets of calibration criteria are provided, one for broad-based equity indices of non-Asian developed economies, and one for small capitalization equity indices. Guidance is also provided for indices that do not fall into these two categories.

360 Albert Street, Suite 1740, Ottawa ON K1R 7X7
613.236.8196 613.233.4552
secretariat@actuaries.ca / secretariat@actuaires.ca
actuaries.ca / actuaires.ca

In accordance with the Institute's Policy on Due Process for the Adoption of Guidance Material Other than Standards of Practice, this research paper has been prepared by the Committee on Life Insurance Financial Reporting, and has received the approval for distribution from the Practice Council on January 19, 2012.

The members of the Designated Group are:

Jas Bhatia
Alexis Gerbeau (Chair)
Lynn Guo
Martin Labelle
Daniel Mayost (OSFI representative)
Ricardo Mitchell
Christian-Marc Panneton
Steven Prince
Jim Snell
Anthony Vaz
Chong Zheng

If you have any questions or comments regarding this research paper, please contact Alexis Gerbeau, Chair, Designated Group, at his CIA Online Directory address, alexis.gerbeau@standardlife.ca.

PR, EG, AG

TABLE OF CONTENTS

1 PURPOSE 5
2 SUMMARY 6
3 THE DATA 7
3.1 Choice of Indices 7
3.2 Choice of Historical Periods 7
3.3 Source of Data and Manipulations 8
4 THE METHOD 8
4.1 Choice of Models 10
4.2 Bootstrap: Choice of Block Sizes 11
4.3 Note on Empirical Percentiles 11
4.4 Criteria for Non-Canadian Indices 12
4.5 Discussion of the Approach Chosen and Alternatives Rejected 12
5 RESULTS 14
5.1 Goodness of Fit 15
5.2 Validation 18
6 THE CRITERIA 19
6.1 Left Tail Criteria 19
6.2 Mean and Volatility Criteria 23
6.3 Right Tail Criteria 23
7 THE APPLICATION OF CRITERIA 24
7.1 Process 24
8 REFERENCES 26
APPENDIX A: MODELS 28
APPENDIX B: BOOTSTRAP APPROACH 30
APPENDIX C: RESULTS 31
Left Tail of Distributions 31
Mean, volatility and Sharpe ratio 32

1 PURPOSE

The existing calibration criteria for equity returns for the valuation of segregated fund guarantees came into effect in 2002 and were established based on the Canadian market data from 1956 to 1999. These criteria, and the methodology used to develop them, were published in 2002 in the report of the CIA Task Force on Segregated Fund Investment Guarantees.

Several factors have contributed to the need for revised criteria:

- An additional 10 years of data since 1999 is now available;
- The dramatic market events of 2008–2009 are expected to have a significant impact on making the criteria more stringent;
- The segregated fund product market has evolved with a trend toward more longer-term products and additional reset features;
- The increasing use of hedging by companies has raised additional issues with regard to calibration of models.

These points were brought out in the report of the Task Force on Segregated Fund Liability and Capital Methodologies, published in 2010. That report recommended that a working group be formed to address the following issues with respect to calibration criteria:

- Review the current criteria and consider adding new criteria for the right tail of the distribution;
- Review the horizons over which the criteria are provided;
- Provide guidance for the projection of future realized volatility of equity returns;
- Consider the relevance of reflecting the volatility regime prevailing at the valuation date in projecting equity returns;
- Determine what is the appropriate time step to use for estimating volatility in the context of hedging; and
- Consider the relationship between insurance contract liabilities calculated with and without the reflection of hedging.

Those recommendations led to the creation of a working group reporting to CLIFR on the calibration of equity returns. As the program outlined above was ambitious and required lengthy work, and because it was felt that a review of the existing criteria of equity returns was a pressing issue, it has been decided to break up the work of this group into two phases. This research paper presents the results of phase I, which covers the calibration of the left and right tails of equity returns. Phase II will cover the calibration of the volatility of equity returns in the context of hedging. Phase II is expected to be completed in 2012.

The scope of this research paper includes:

- Update of the existing criteria for the left tail of Canadian equity returns at percentiles 2.5%, 5% an 10%, for the one-, five- and 10-year horizons;
- Addition of criteria for the right tail of equity returns;

- Addition of criteria for the 20-year horizon; and
- Addition of criteria for non-Canadian equity returns.

Note that establishing criteria for returns on fixed-income funds is out of the scope of this paper. Guidance for fixed-income returns is expected to be provided by this working group in 2012.

2 SUMMARY

Two sets of calibration criteria have been established for the left tail of equity return distributions: one applicable to broad-based indices of developed non-Asian economies (L1 indices), and one applicable to small capitalization indices (L2 indices).

Calibration criteria have been set for the 2.5th, 5th and 10th percentiles of the accumulation factors for the one-, five-, 10- and 20-year horizons. The following table presents the maximum values for the accumulation factors:

Left tail calibration criteria	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
L1 indices	0.74	0.81	0.88	0.70	0.80	0.95	0.80	0.95	1.20	1.25	1.65	2.25
L2 indices	0.68	0.76	0.85	0.60	0.70	0.90	0.70	0.90	1.20	1.10	1.55	2.35

In addition, constraints have been established for the volatility and the mean of equity returns.

The minimum value for the standard deviation of the one-year accumulation factor for U.S. broad-based indices is 16.5%. For all other L1 indices, the minimum value for the standard deviation is 17.5%. For L2 indices, the minimum value for the standard deviation is 23%.

The range for the expectation of the one-year accumulation factor is 8% to 12% for all L1 indices, and 11% to 15% for L2 indices.

These criteria are summarized in the following table:

Mean and volatility criteria	Minimum for the mean	Maximum for the mean	Minimum for the volatility
U.S. broad-based indices	8%	12%	16.5%
L1 indices other than U.S. indices	8%	12%	17.5%
L2 indices	11%	15%	23%

Finally, calibration criteria for the right tail percentiles relative to the median have been established for the one-year horizon at the 90th, 95th and 97.5th percentiles. These criteria are met if the differences between the right tail percentiles and the median are greater or equal to the criteria. These criteria apply to the L1 and L2 indices.

Right tail calibration criteria	1-year		
	90th	95th	97.5th
L1 and L2 indices	0.18	0.24	0.30

The criteria have been determined using monthly total returns from January 1956 to June 2010. The data for the TSX, S&P 500, FTSE, MSCI EAFE and Russell 2000 have been considered. Results for a range of stochastic models have been considered, as well as for the bootstrap approach.

Details on the development of the criteria are provided in section 6.

The Application of the Criteria

The criteria apply to the scenarios generated for the valuation. This represents a change from the current approach where it is the model that is subject to the constraints. If a closed-form formula exists for a statistic subject to the criteria, it would be sufficient to test that the theoretical value of the statistic calculated using the closed-form formula meets the criteria, as long as a large number of scenarios is used for valuation, and the actuary tests that the discrepancy between the theoretical value and the value calculated with the scenario set is not material.

The calibration process also provides rules for indices other than L1 and L2 indices.

Details on the process of applying criteria are provided in section 7.

3 THE DATA

3.1 Choice of Indices

For the purpose of this research paper, we attempted to choose indices that are commonly used when modeling benchmarks in the valuation of segregated fund investment guarantees.

We reviewed and analyzed the data of large capitalization indices in the following countries/regions:

Indices by Region	
Countries/Regions	**Name of Index**
Canada	S&P TSX Composite
United States	S&P 500
United Kingdom	FTSE All-Share
Europe, Australasia and Far East	MSCI EAFE

A detailed description of the indices mentioned above is also available in appendix A of the educational note Investment Return Assumptions for Non-Fixed Income Assets for Life Insurers.

In addition, we reviewed the data of the Russell 2000, which is the most representative small capitalization index in the United States.

3.2 Choice of Historical Periods

The 2002 CIA Task Force on Segregated Fund Investment Guarantees used monthly total return data for the TSE300 for the period from January 1956 to December 1999. In order to ensure continuity with the previous task force, we expanded the range to include monthly total return data up to June 2010 for the S&P TSX Composite. For other indices, we attempted to use similar ranges when the data were available. Total returns monthly data were used in all cases. This choice of historical period is also similar to the historical period (1955 to 2005) used for the work conducted by the American Academy of Actuaries on variable annuities in the United States.

In addition, we performed analysis using an alternate period for the U.S. market by using monthly total return data from December 1927 to June 2010.

The table below summarizes the historical period for which we performed our analysis

Historical Period by Index		
	Period	
Name of Index	**From**	**To**
S&P TSX Composite	January 1956	June 2010
S&P 500	January 1956	June 2010
S&P 500	December 1927	June 2010
Russell 2000	January 1956	June 2010
FTSE All-Share	January 1956	June 2010
MCSI EAFE	December 1969	June 2010

3.3 Source of Data and Manipulations

We used Bloomberg as our main source of the total return monthly data. In some cases, the data series were incomplete and we had to complete the data series from other sources. This section describes how the various data series were completed.

- For the S&P TSX Composite index, we used the Bloomberg data for the period from January 2000 to June 2010. Prior to January 2000, we used the data series published by the CIA in the previous task force.
- For the S&P 500 index, we used the Bloomberg data for the period from February 1988 to June 2010. Prior to February 1988, we used the data series published by Ibbotson for U.S. total return on large capitalization companies.
- For the Russell 2000 index, we used the Bloomberg data for the period from December 1978 to June 2010. Prior to December 1978, we used the data series published by Ibbotson for U.S. total return on small capitalization companies.
- For the FTSE All-Share index, we used the Bloomberg data for the period from December 1985 to June 2010. Prior to December 1985, we used the data series published by Global Index for U.K. total return on large capitalization companies.
- For the MSCI EAFE index, we only used the Bloomberg data.

In order to ensure consistency for the above indices, we compared the Bloomberg data to the other sources of data available for the overlapping periods. This validation was conclusive.

4 THE METHOD

Two broad families of methods have been applied for deriving calibration criteria: the model-based approach and the bootstrap (model-free) approach. In addition, empirical percentiles for the one-year horizon were considered.

Model-based Approach

The model-based approach consists of selecting a stochastic parametric model for equity returns, then fitting the model to data (estimating the parameters) using a statistical method. The required percentiles of equity returns are derived from that model, either with a closed-form formula when available or with Monte Carlo simulations. This approach therefore relies on the appropriateness of a particular choice of model. Even though statistical measures of goodness of fit exist for selecting models, no model can be viewed as the definitively best model to be used in all circumstances. Instead, a variety of valid models exist, some of which reproduce some stylized facts of historical returns better than others. Therefore, the model-based approach involves a certain level of subjectivity in the choice of the model form.

Bootstrap Approach

Bootstrapping is a well-established non-parametric approach to estimating properties of an estimator. In our context, the bootstrap approach consists of generating a sample of equity return paths by randomly picking historical blocks of returns, where all blocks are of the same size. A block of returns of size $\boldsymbol{n}$ is any series of $\boldsymbol{n}$ consecutive monthly returns from the historical dataset, i.e., returns from month $\boldsymbol{i}$ to month $\boldsymbol{i}+\boldsymbol{n}-1$, for any $\boldsymbol{i}$. Note that these blocks are possibly overlapping. The percentiles of equity returns are then calculated from the sample. Details on the bootstrap approach are provided in appendix B. Section 4.2 discusses the choice of block size.

The bootstrap approach is arguably less subjective than the model-based approach, as it does not depend on the choice of a particular model. The selection of the block size remains a choice of the modeller though. Also, a weakness of the bootstrap approach is that it will never generate returns more extreme that the actual historical record.

The approach chosen for this paper is to apply both the model-based method with a number of models, and the bootstrap method with various block sizes. Criteria were established by considering the range of results obtained. This can be seen as a balanced solution, where we consider many approaches rather than selecting a single method to be preferable to all others. We believe that this approach helps to reduce the subjectivity of the exercise.

We end this section by noting that the existing 2002 criteria were derived using the model-based approach. The models considered in 2002 are listed in the following section.

4.1 Choice of Models

CIA Task Force Report (2002)

In deriving the criteria specified in the 2002 CIA Task Force report, the then task force considered the following models:

- Two-factor regime-switching lognormal model;
- Stable distribution model; and
- Stochastic volatility lognormal model.

Based on this fit, a set of calibration points were selected that overall represented the results of the three fitted models. The calibration points were derived by ensuring that they would be met by all of these models when appropriately calibrated.

American Academy of Actuaries (AAA) Report (2005)

In deriving the criteria specified in the 2005 AAA report, the then task force considered the following models:

- Independent lognormal model;
- Two-factor regime-switching lognormal model;
- Three-factor regime-switching lognormal model (fitted to daily data);
- Three-factor regime-switching lognormal model (fitted to monthly data); and
- Stochastic log volatility model with varying drift.

The stochastic log volatility model with varying drift was used for the development of the “preliminary” calibration points. This model was not necessarily preferred above the others, but was chosen as it captured many of the important dynamics of short-term equity returns, such as negative skewness, positive kurtosis and clustered volatility. The preliminary calibration points were then adjusted modestly so that a range of common (but suitably calibrated) models would pass the standard. For the pros and cons of each of the models, please refer to the AAA report.

Our Approach

Our task force investigated the following models:

- Lognormal model (LN);
- Two-factor regime-switching lognormal (RS2LN) model;
- Three-factor regime-switching lognormal (RS3LN) model;
- Two-factor regime-switching draw down (RS2DD1) model;
- Stochastic volatility lognormal (SVL) model; and
- Correlated stochastic volatility lognormal (CSVL) model.

The RS3LN model was subsequently withdrawn. When that model was fitted to the S&P 500 data using maximum-likelihood estimators, the model naturally reverted back to a RS2LN model (i.e., the probability of having a third regime was 0). For the indices where the model provided a third regime with non-zero probability, the marginal increase in likelihood value (relative to the two-factor version of the model) was minimal. Twelve parameters need to be estimated—as a result of which they became more volatile. Consistent with the AAA report, this task force found the model to be more difficult to use and calibrate than the two-factor model yet it did not produce a meaningful improvement in fit to the data.

All remaining models were calibrated using maximum likelihood estimators (MLE). Two different versions of the SVL model were investigated. The first version (SVL) assumes that there is no correlation between returns and the volatility of returns. This model produced unrealistic results. The second version (CSVL) assumes a negative correlation between returns and the volatility of returns. We retained the second version because it provided a better fit. The non-linear filtering methods of Clements and White and of Watanabe were used to calibrate these models.

4.2 Bootstrap: Choice of Block Sizes

With time series data, simply drawing individual observations with replacement is not sufficient to replicate the underlying distribution of returns, because of possible serial correlation within the data. To preserve this serial correlation, blocks of data were chosen rather than individual values. The optimal block size is not evident.

- If the blocks are too short, the serial correlation within the original data is lost—if there is positive serial correlation, then the results will exhibit thinner tails than the actual distribution as the bootstrapping will not capture the prolonged negative returns.
- If the observations are negatively serially correlated, then choosing small blocks in bootstrapping will result in tails that are too thick, as it will not capture the recoveries in the market after a fall.
- If the observations are serially independent, the choice of block sizes should not matter.

The optimal block size should be chosen as the smallest block size for which the serial correlation between the observations has been largely removed for any larger block size. Unfortunately, our analysis showed that some negative correlation persisted for very large block sizes. Due to the limited number of historical data points which limited the block size, we considered three, six and 12 months block size.

4.3 Note on Empirical Percentiles

In calculating the historical empirical percentiles, we had to handle the starting point. Typically to avoid bias, non-overlapping periods would be selected. For example, with five-year periods, starting on January 1956, the choices are: January 1956 to January 1961, January 1961 to January 1966, and so on. Alternatively, if starting on February 1956, the periods would be February 1956 to February 1961, February 1961 to February 1966, and so on. Both sets are composed of non-overlapping periods, but will provide different results as the starting point is different. To remove the bias introduced by selecting the starting point, all possible starting points were considered and the resulting percentiles averaged. For the five-year accumulation factors percentiles, there are 60 possible starting points with monthly data. So we determined the percentiles for each of the 60 different starting points for the five-year non-overlapping periods and then averaged the 60 percentiles to establish the empirical percentiles. It should be noted that with 54 years of data, for the one-year accumulation factor, the 2.5th percentile was determined by linear interpolation between the two lowest observations: the lowest observation is the 1.82th (1/55) percentile and the second lowest observation is the 3.64th (2/55) percentile.

4.4 Criteria for Non-Canadian Indices

A significant portion of segregated fund exposure for Canadian insurers is related to non-Canadian markets. A number of suitable indices for non-Canadian equity markets have been identified based on the breadth of their coverage, their class and characteristics, and on their importance for Canadian insurers. The historical data for the TSX, S&P 500, FTSE, Russell 2000, Topix, MSCI EAFE and Hang Seng have been considered for developing the calibration criteria. A general description of these indices can be found in the educational note Investment Return Assumptions for Non-Fixed Income Assets for Life Insurers published in March 2011.

In order to keep the number of criteria small, a common set of criteria would be established for a group of indices sharing similar statistics. Other factors such as diversification, geography, size of capitalization, and the relative importance of an index for Canadian insurers were considered.

We did not want to assume that any one market would consistently outperform others or present a superior risk-return profile over the long term. Judgment was therefore applied for developing guidance for indices that show very high historical returns.

The calibration criteria apply to the accumulation factors of non-Canadian indices in local currency. When non-Canadian indices are modeled, the foreign exchange rate would also be considered.

4.5 Discussion of the Approach Chosen and Alternatives Rejected

Absolute Returns vs. Equity Risk Premium

The working group decided to consider historical equity returns in absolute terms, as opposed to excess equity return relative to risk-free interest rates. This implies that the level of risk-free interest rates prevailing at the valuation date does not impose additional constraints on the projected equity returns.

As with many aspects of this project, there was no absolute right or wrong approach, nor was complete consensus achieved within the working group on this point.

Economic theory suggests a positive relationship between the level of interest rates and future expected equity returns. As interest rates have been at a much higher level on average over the period spanned by the equity data than they have been in recent years, one could argue that it would be more conservative to use prevailing risk-free interest rates combined with criteria about equity risk premiums. However, the empirical relationship between interest rates and equity returns is much less clear than what the theory predicts.

Some members also argue that if the expected equity return decreases with the level of interest rates, the volatility of equity returns should in turn also decrease, on the basis that the relationship between expected return and volatility should be maintained. By that theory, the impact on the left tail of equity returns of low interest rates would be mitigated by a decrease in volatility. Other members believe that it is the price of risk, as measured by the Sharpe ratio, that can be expected to be stable. According to this argument, a decrease in expected equity return following a decrease in interest rates would not imply a decrease in the volatility, i.e., the distribution of equity returns would be shifted to the left, lowering the left tail. There is, therefore, no consensus on this issue. In the absence of empirical evidence, it has been decided to retain the approach of considering historical equity returns adopted in 2002 for setting criteria. This is consistent with the standards of practice for setting the equity return assumption in the context of a deterministic application of CALM, as described in the educational note Investment Return Assumptions for Non-Fixed Income Assets for Life Insurers published in March 2011.

Mean Reversion

All the models considered under the model-based approach with the exception of the RSDD model assume the absence of a dependence structure in the level of returns over long periods of time. In other words, these models assume that a recent episode of very high or low returns provides no information for forecasting returns in following years. The bootstrap approach also implicitly assumes the independence of returns over periods of time extending beyond the block size. Many actuaries believe that this assumption is not well founded, and that markets do exhibit mean reversion, i.e., that periods of low returns tend to be followed by periods of high returns, and vice versa.

Those sharing this view would argue that ignoring mean reversion provides a layer of conservatism in our calibration criteria.

The issue of mean reversion, or more generally of the dependence structure of returns in time, is controversial and has been the subject of much econometric research. We have decided to exclude mean reversion in establishing calibration criteria for two reasons. First, as mentioned above, there is no consensus on this issue in the academic or actuarial communities. Secondly, the percentiles of equity returns over long-term horizons are very sensitive to the strength of mean reversion assumed and the form of dependence structure chosen. Therefore, allowing for some temporal dependence structure would render the exercise of setting calibration criteria very judgmental. We preferred the simpler approach of assuming no dependence, acknowledging that this can be viewed as providing an element of conservatism.

Choice of Historical Period

Another source of subjectivity is the choice of the historical period, in particular the decision of whether or not to discard the pre-1956 data, and therefore the experience of the 1930s Great Depression. We will come back to this issue in section 6.1.

Making Choices is Unavoidable

The exercise of establishing criteria does not rely only on statistical analysis, and involves a certain level of professional judgment. The impact of the modeler choices is especially important for long-term criteria, where any difference in approach or assumptions compounds and results in large differences in projected percentiles.

Consistency with Prior Approaches

An additional consideration in our choices was the desire not to introduce a radical change in approach from the method used in 2002 for developing the existing criteria. The coming of IFRS 4 in the future is likely to render our real-world criteria obsolete. We felt that an update of the 2002 analysis was more appropriate than a revolutionary change in approach.

Considering all elements discussed above, the working group believes that our approach is sensible and balanced.

5 RESULTS

Using the four models selected (LN, RS2LN, RS2DD1 and CSVL), we generated random paths of returns over different horizons (one-, five-, 10- and 20-year). To minimize sample error, one million paths were generated. The resulting 2.5^{th}, 5^{th} and 10^{th} percentiles are presented in appendix C. For these models, the expected return over a one-year holding period was calculated with its standard deviation. The Sharpe ratio was obtained from these numbers assuming a 4% risk-free rate.

In addition to the four models selected, we also generated random paths of returns with the bootstrap technique. We used block size of three, six and 12 months. The percentiles for this approach are also presented in appendix C.

For the TSX, left tail percentiles are lower than current calibration criteria at all horizons for the RS2LN and CSVL models. This is also true, but to a lower extent, for the bootstrap method. As mentioned in the introduction, this is not surprising given the poor market returns experienced in 2008 and 2009.

The percentiles for the S&P 500 produced by the RS2LN and CSVL models are very similar to those for the TSX. Those for the FTSE are also similar up to the five-year horizon, and are higher for longer horizons, because of the higher mean of the FTSE. Percentiles for the MSCI EAFE are lower than those for the TSX, S&P 500 and FTSE, except for the CSVL model. However, the percentiles for the MSCI EAFE are derived using a different historical period. Further analysis showed that the differences diminish when we use the same period.

The bootstrap method produces percentiles similar to those for the RS2LN and CSVL models for the one-year horizon, but generally higher percentiles over longer horizons, and the difference increase with the length of the horizon.

As expected, percentiles for the RS2DD1 model over long horizons are significantly higher than those obtained with other methods, because of the mean reversion assumed in this model. The differences increase with the length of the horizon. The LN model produces the second larger percentiles. An exception to these general rules is the percentiles obtained with the CSVL model for the MSCI EAFE, which are out of proportion with those produced by other methods.

There is no clear relation between percentiles produced by the RS2LN and CSVL models. The RS2LN model produces lower percentiles than the CSVL model for the TSX, higher percentiles than the CSVL model for the S&P, and both models produce similar percentiles for the FTSE.

For many models and indices, the 2.5th and 5th percentiles are lower at the five-year horizon than at the one-year horizon, and then increase at the 10-year and 20-year horizons. This pattern is explained by the combination of two opposing forces: the positive expected return raising the expected accumulation value over time, and the volatility of returns widening the dispersion of outcomes over time, and therefore lowering low percentiles. Low percentiles tend to decrease over the first few years of projection due to the impact of volatility, and then increase beyond a certain point as the impact of the positive expected return more than offsets the impact of volatility.

As noted above, the working group decided to work with several models in parallel because there was no consensus on whether any one of these models is necessarily the best in all cases. Nor is there necessarily consistency in which model produces the highest or lowest results when applied to a range of different time series and indices.

5.1 Goodness of Fit

We determined parameters for the four models we selected using maximum likelihood estimation. The models are described in appendix A and reference material is provided for the details of the parameter estimation methodology used.

The following tables present the log-likelihood value for each of the models and each of the five indices we considered. We computed the Akaike information criteria (AIC) and the Schwartz Bayes criteria (SBC). The interpretation of these criteria is that given a set of plausible models for the data, the preferred model is the one with the highest criteria value. When estimating parameters using maximum likelihood estimation, it is possible to increase the likelihood by adding parameters: as the RS2LN model is a special case of the more general RS2DD1 model, the RS2DD1 model will always provide a log-likelihood equal or higher to the RS2LN model. To compensate for this, both the AIC and the SBC include a penalty which increases with the number of parameters in the model. This penalty discourages overfitting. The penalty is larger under the SBC than under the AIC.

Jan 1956–Jun 2010		**TSX**			**S&P 500**		
		Nb data points		**653**	**Nb data points**		**653**
Model	**Nb. Param.**	**Log-lik.**	**AIC**	**SBC**	**Log-lik.**	**AIC**	**SBC**
LN	2	1090.0	1088.0	1083.5	1129.8	1127.8	1123.4
RS2LN	6	1139.7	1133.7	1120.2	1166.2	1160.2	1146.8
RS2DD1	8	1144.0	1136.0	1118.0	1173.8	1165.8	1147.9
CSVL	5	1147.6	1142.6	1131.4	1188.6	1183.6	1172.4

Jan 1956–Jun 2010		Russell 2000			FTSE		
		Nb data points		653	Nb data points		653
Model	Nb. Param.	Log-lik.	AIC	SBC	Log-lik.	AIC	SBC
LN	2	913.9	911.9	907.4	991.3	989.3	984.8
RS2LN	6	960.2	954.2	940.8	1066.4	1060.4	1046.9
RS2DD1	8	969.9	961.9	944.0	1069.8	1061.8	1043.9
CSVL	5	982.7	977.7	966.5	1091.7	1086.7	1075.5

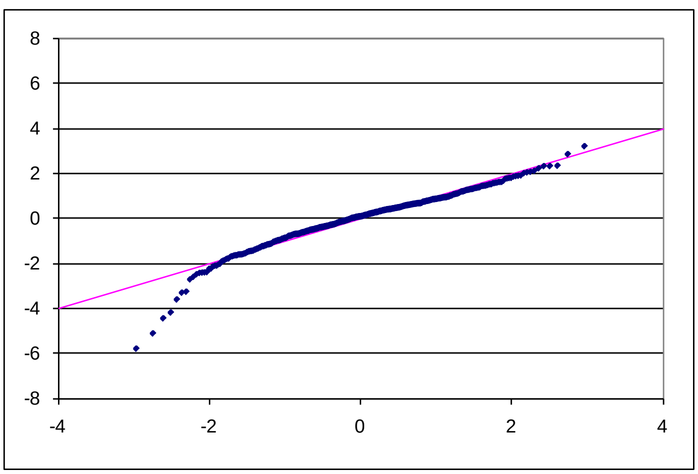
Dec 1969–Jun 2010		EAFE		
		Nb data points		486
Model	Nb. Param.	Log-lik.	AIC	SBC
LN	2	840.5	838.5	834.3
RS2LN	6	889.5	883.5	870.9
RS2DD1	8	891.0	883.0	866.3
CSVL	5	890.6	885.6	875.1

It is interesting to note that for all seven indices, the CSVL model provides the best fit under both the AIC and the SBC criteria.

However, these numbers do not tell everything about the distribution of returns. So we have also looked at the QQ plots of the residuals under the lognormal model to get a better sense of the tail fatness.

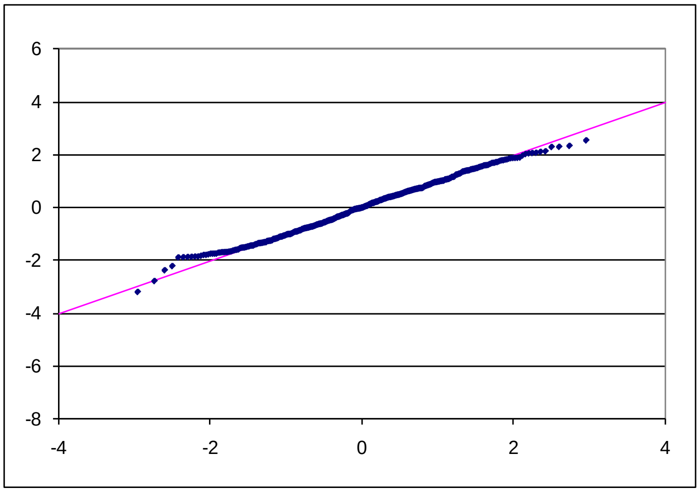
Graph 1 shows the QQ plot of the residuals under the lognormal model for the TSX index. The graph clearly shows that the left tail of the distribution presents tail fatness: much lower (negative) returns have occurred in the past compared to what would have been expected under a normal distribution. However, the right tail does not present any significant tail fatness, meaning that returns that have occurred in the past were not higher than those that would have been expected under a normal distribution. This indicates asymmetry in the distribution of returns.

Graph 1



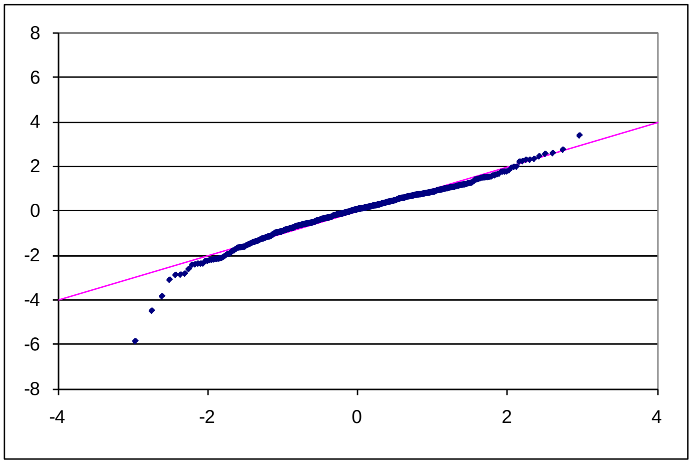
Graph 2 shows the QQ plot of the residuals under the RS2LN model for the TSX index. We observe a much better fit of the left tail, but a slight deterioration on the right tail.

Graph 2



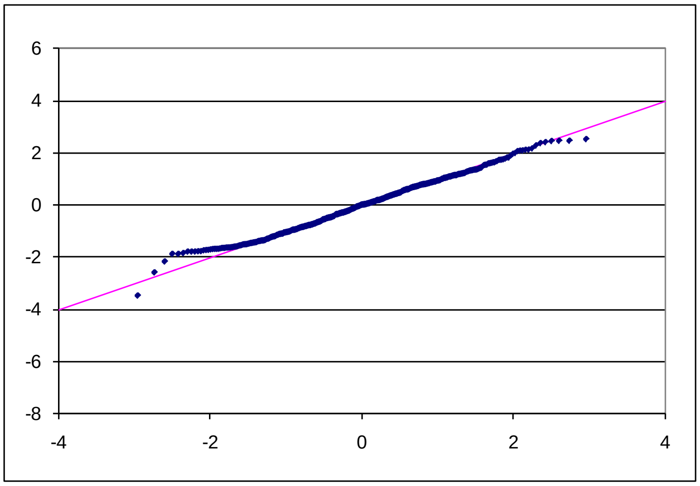
Graph 3 shows the QQ plot of the residuals under the lognormal model for the S&P 500 index. The same observations as for the TSX apply to the S&P 500.

Graph 3



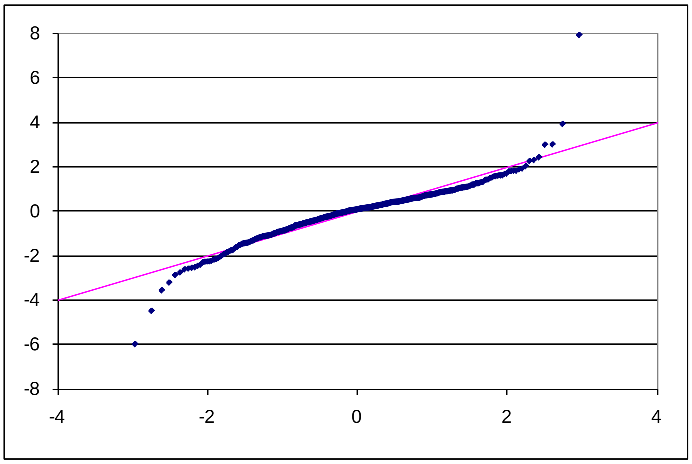
Graph 4 shows the QQ plot of the residuals under the RS2LN model for the S&P 500 index. Again, we observe a better fit of the left tail, and no material change on the right tail.

Graph 4



Graph 5 shows the QQ plot of the residuals under the lognormal model for the FTSE index. This is the only index to show some right tail fatness. In particular, one observation is especially high: for the month of January 1975, the log-return of the FTSE was 43.2%. For comparison, for that month, the TSX provided a return of 15.3% and the S&P 500 a return of 11.8%.

Graph 5



5.2 Validation

We did the following to validate the model and the calibration:

- The RS2LN model was recalibrated to the 1956–1999 data (same period as used by the 2002 working group). Their results were replicated, with the exception of some small differences due to data inconsistencies.
- The models were recalibrated independently by different members of the working group, using calibration tools that were independently built. The same parameters are achieved.

6 THE CRITERIA

6.1 Left Tail Criteria

Number of Calibration Sets

Given the similarity of left tail percentiles for the TSX, S&P 500, FTSE and MSCI EAFE, a single set of calibration criteria was established for all broad-based indices of developed economies with the exception of East Asian economies. These economies are henceforth referred to as the developed non-Asian economies. The East Asian economies were excluded because of significant differences in historical market experience. Broad-based indices are indices whose purpose is to reveal the performance of an entire market. They generally cover a large proportion of the capitalization of a market.

The decision to have a single set of criteria for these economies is also supported by the view that, prospectively, there is no compelling reason to believe that one developed non-Asian economy will outperform others. Finally, having a single set of criteria was considered more practical.

We define developed economies as those considered as developed by MSCI in their index development. MSCI classifies countries into three categories: developed, emerging and frontier. The three main criteria used by MSCI to classify economies are economic development, size and liquidity of markets, and market accessibility. Details are provided in the document MSCI Market Classification Framework.

MSCI currently considers the following markets as being developed:

Developed Countries			
Americas	**Europe and Middle East**		**Pacific**
Canada	Austria	Italy	Australia
United States	Belgium	Netherlands	Hong Kong
	Denmark	Norway	Japan
	Finland	Portugal	New Zealand
	France	Spain	Singapore
	Germany	Sweden	
	Greece	Switzerland	
	Ireland	United Kingdom	
	Israel		

The developed non-Asian economies are therefore the ones included in the table above, with the exception of Hong Kong, Japan and Singapore. We henceforth refer to broad-based indices of developed non-Asian economies as L1 indices.

Some economies considered as developed by MSCI are of small size, such as Greece and Belgium. A fund tied to one or a small number of these small developed economies should be subject to more stringent criteria than a fund tied to a well-diversified basket of developed economies or a large developed economy such as the United Kingdom or Germany. Therefore, the indices to which a single set of criteria will apply are those that are composed of a diversified basket of L1 indices or of an L1 index of a large economy. Indices composed of a small number of L1 indices of small economies will be subject to a different treatment (they will fall under case 2 in the decision tree presented in section 7.1).

A second set of left tail calibration criteria has been developed for the small capitalization indices, based on results for the Russell 2000 index. Having a separate set of criteria is justified by the higher volatility and lower left tail percentiles of this index. It is also generally accepted that small capitalization indices tend to be more volatile than broad-based indices. We henceforth refer to small capitalization indices as L2 indices.

In developing the criteria L1 indices, more weight was assigned to the TSX and S&P 500 results, because of the important exposure of Canadian companies to these two indices, and because the S&P 500 is the largest, most diversified, and arguably most representative broad-based index in the world.

Models Considered

Left tail criteria have been set primarily by considering the results produced by the RS2LN and CSVL models. The LN model does not produce enough fat tails but has been included as a benchmark. The RS2DD1 model assumes mean reversion of returns and was presented for illustration purposes. As the RS2LN and CSVL models result in lower left tail percentiles than the bootstrap approach in most cases, especially for long horizons, this choice is considered prudent.

Left tail criteria are greater or equal to the largest of the percentiles produced by the RS2LN and CSVL models for the TSX and S&P at all horizons. This provides some leeway to the actuary for the choice of a model.

The working group felt that it would not be useful to align calibration criteria to the results produced by a single model. Doing so could have resulted in that model becoming the default model used by actuaries, which was not an objective of the working group. Also, calibration criteria represent upper limits for models. It is expected that actuaries would fit their models to data using theoretically sound statistical methods, which could result in many circumstances in more conservative percentiles than the calibration criteria.

Finally, we felt that this assignment did not require excessive conservatism in the setting of calibration criteria. Conservatism in the calculation of insurance contract liabilities for segregated funds is provided by the use of a high level of CTE, and not in a conservative parameterisation of the stochastic model for investment returns.

Choice of the Historical Period

Left tail criteria have been set using the data from 1956 to June 2010. Various factors were considered for discarding the pre-1956 data.

Great Depression not likely to be repeated. It is generally accepted that the Great Depression was caused, at least in part, by inept monetary policies. The proponents of this thesis generally believe that nowadays central banks have a better understanding of the impact of monetary policies and have developed better management tools.

The financial crisis experienced in 2008–2009 is sometimes presented as an argument in favour of considering the market experience of the 1930s as being relevant for forecasting future equity returns. The members of this working group do not share this view, and are of the opinion that the market experience since 2008 and the experience of the 1930s are not comparable. The stock market did not come back to its level prevailing prior to the 1929 crash until 1944, a period of 15 years. Also, the volatility of the market following the 1929 crash remained at an abnormally high level until the mid-1930s. In contrast, the stock market following the 2008 crisis recovered most of its losses within two years, as of writing this report.

Other elements of conservatism. Discarding the pre-1956 data can be viewed as a choice that counterbalances some elements of conservatism in our approach. These conservative elements are (a) the assumption of the absence of any mean reversion in returns, and (b) the reliance on the RS2LN and CSVL models instead of the bootstrap approach for deriving criteria for long-term horizons. As discussed in section 4.5, establishing criteria necessarily involves applying some judgment. Modeling choices must be considered in their entirety, as the purpose is to have a balanced approach as a whole.

Results shown in appendix C show that the two elements of conservatism mentioned above counterbalance to a great extent the choice of discarding the pre-1956 data. Results for the 2.5th percentile at the 20-year horizon for the S&P 500 are reproduced in the table below. The 2.5th percentiles for the bootstrap approach, the RS2LN and RS2DD1 models are significantly lower when using the data from 1926. However, the percentiles for the RS2LN model are significantly lower than for the bootstrap approach and the RS2DD1 model, for both historical periods. As a result, the percentile of the RS2LN model using the data from 1956 (1.16), on which the criterion is based, is not significantly higher than that of the bootstrap approach (ranging from 0.92 to 1.07), and is lower than that of the RS2DD1 model (1.25), when the data from 1926 are used. The same relationship between models and historical periods holds for other horizons.

2.5th percentile at the 20-year horizon for the S&P 500		
Model	Jan 1926 to June 2010	Jan 1956 to June 2010
Bootstrap 3-month	1.07	1.52
Bootstrap 6-month	1.05	1.41
Bootstrap 12-month	0.92	1.37
RS2LN	0.55	1.16
RS2DD1	1.25	2.15

Consistency with 2002 approach and upcoming IFRS. As mentioned in section 4.5, our choice was also motivated by the desire of not introducing a radical change in the context of the coming of IFRS 4 in the future. The consistency with the approach taken in 2002 was therefore preferred.

Lack of non-U.S. data prior to 1956. Finally, the fact that the only available data prior to 1956 were for the U.S. economy was considered a disadvantage. The working group felt there was merit in the consistency of using the same time periods across several economies, which would not be possible if pre-1956 data were used.

Other Considerations

Calibration criteria for horizons longer than one year have been rounded up to the nearest 0.05. An additional layer has been added to the criteria for the 20-year horizon. Because of the important uncertainty in estimating left tail percentiles for such a long horizon, we wanted to avoid the 20-year calibration criteria becoming the binding constraint for the calibration process. The important uncertainty in estimating percentiles for long horizons is revealed by the standard deviation of the percentiles estimators. The next table shows the 2.5th percentile estimates and their standard deviation using the RS2LN model for the TSX.

TSX (January 1956 to June 2010)—RS2LN model		
Horizons	2.5th percentile	Standard deviation
1-year	0.71	0.06
5-year	0.61	0.15
10-year	0.67	0.29
20-year	0.99	0.74

Finally, as these criteria will be applicable to the actual scenarios used for valuation, and because the practice is to update the parameterization of models on an annual basis using the most recent data, we wanted to reduce the likelihood that the actuary would have to adjust the parameters in a few years from now to satisfy criteria.

The Left Tail Criteria

The following table presents the maximum values for the left tail of accumulation factors for the one-, five-, 10- and 20-year horizons.

Left tail calibration criteria	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
L1 indices	0.74	0.81	0.88	0.70	0.80	0.95	0.80	0.95	1.20	1.25	1.65	2.25
L2 indices	0.68	0.76	0.85	0.60	0.70	0.90	0.70	0.90	1.20	1.10	1.55	2.35

In comparison to the criteria for L1 indices, the criteria for the 10th percentile for the L2 indices are equal at the 10-year horizon, and even larger at the 20-year horizon. This is explained by the higher mean of small capitalization indices, which more than offset its higher volatility for long horizons.

The following table shows the existing calibration criteria for the TSX.

Existing left tail calibration criteria	1-year			5-year			10-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
TSX	0.76	0.82	0.90	0.75	0.85	1.05	0.85	1.05	1.35

The new criteria for L1 indices over the one-year horizon are slightly lower than the existing criteria for the TSX. This was expected because of the addition of the data for the 2008–2009 financial crisis. Decreases in the criteria for the five- and 10-year horizons are more important, due to the compounding effect.

6.2 Mean and Volatility Criteria

Despite the fact that a single set of left tail criteria has been established for L1 indices, distinct constraints for volatilities have been established, because of differences in historical volatilities. For U.S. broad-based indices, the minimum value for the standard deviation for the one-year accumulation factor is 16.5%. For all other L1 indices, the minimum value for the standard deviation is 17.5%. The lower constraint for U.S. broad-based indices is justified by a lower historical volatility for the S&P 500 and by the fact that the U.S. economy is more diversified.

For L2 indices, the minimum value for the standard deviation for the one-year accumulation factor is 23%. This number is close to the historical standard deviation of the Russell 2000 index.

For all L1 indices, including the S&P 500, the range for the expectation of the one-year accumulation factor is 8% to 12%. In order to allow the actuary to apply judgment in setting expected return, the lower bound of that range has been set at a level that is lower than historical average returns for the broad-based indices covered in this study.

For L2 indices, the range for the expectation of the one-year accumulation factor is 11% to 15%. This range has been obtained by shifting the range for L1 indices upward by 3%, which is close to the differential in historical means between the Russell 2000 index and broad-based indices.

These criteria are summarized in the following table.

Mean and volatility criteria	Minimum for the mean	Maximum for the mean	Minimum for the volatility
U.S. broad-based indices	8%	12%	16.5%
L1 indices other than U.S. indices	8%	12%	17.5%
L2 indices	11%	15%	23%

6.3 Right Tail Criteria

The Task Force has added right tail criteria to the requirements. These right tail criteria have been added to capture the increasing risk caused by ratchet/reset and lock-in features that have become more common in products now being sold. To the extent some funds have occasional upward jumps in value, this jump can significantly increase the guarantee value if a policyholder elects to lock-in that value or the lock-in is contractually automatic.

Calibration criteria for the right tail percentiles relative to the median have been established for the one-year horizon at the 90^{th}, 95^{th} and 97.5^{th} percentiles. These criteria are met if the differences between the right tail percentiles and the median are greater or equal to the criteria. These criteria apply to the L1 and L2 indices.

Right tail calibration criteria	1-year		
	90^{th}	95^{th}	97.5^{th}
L1 and L2 indices	0.18	0.24	0.30

The motivation for establishing right tail criteria in relation to the median is that some actuaries make discretionary downward adjustments in mean returns as an added element of conservatism. Establishing right tail criteria in absolute terms would have constrained the actuary in making downward adjustments. We wanted to avoid this outcome.

The right tail criteria have been derived by considering statistics for all L1 and L2 indices.

7 THE APPLICATION OF CRITERIA

The criteria produced by this working group would be satisfied by the scenarios being used in a valuation. If a closed-form formula exists for a statistic subject to the criteria, it would be sufficient to test that the theoretical value of the statistic calculated using the closed-form formula meets the criteria, as long as a large number of scenarios is used for valuation, and the actuary tests that the discrepancy between the theoretical value and the value calculated with the scenario set is not material.

This approach differs from the approach recommended by the 2002 task force. Under that approach, the test was applied to whether the model chosen met various criteria when that model was fitted to the prescribed data set of TSX returns from 1956 to 1999. Having satisfied this test, the model could then be applied to any index and any time periods.

This working group has decided that the model-test approach is no longer needed. At the time of the 2002 task force, there was considerable debate about which mathematical forms of models would be appropriate. The model-test was attempting to assess whether the mathematical form of the model was appropriate. As of the date of this writing, it is recognized that there are a large number of potentially valid mathematical forms of models. The view of this working group is that the model itself does not need to be validated as long as the results of the model meet the stated criteria.

We believe that this new approach will narrow the range of practice, which is one of the objectives of the CIA. As well, the principle of applying the criteria to scenarios rather than testing the model is consistent with the approaches being developed for interest models.

A disadvantage of having fixed numerical criteria is that these criteria need to be updated more frequently. However, the working group’s view is that it is a relatively routine process to update these criteria from time to time to reflect then-current experience. We believe that updating the criteria every five years would be reasonable.

Because the criteria have been established using data up to June 2010, it is possible that a model fails to meet the criteria when using more recent data, even where an actuary is using one of the models that have been considered in the development of the criteria. This means that an actuary could have to adjust the model parameters obtained from a statistical fitting in the future if the market experience after June 2010 is favourable. The working group’s view is that this does not represent a major disadvantage.

7.1 Process

To model the investment returns of a specific fund, a proxy for the fund would be constructed. The proxy usually takes the form of a linear combination of market indices.

The criteria established in this research paper apply to the investment returns generated for equity indices that are used in the composition of the proxies.

The criteria are to be applied by working through the decision tree described below. The guiding principle here is to use data where such are available and credible.

The decision tree is as follows:

Case 1: If a large proportion of the index is comprised of a diversified basket of L1 indices, of an L1 index of a large economy, or of L2 indices, then the relevant set of calibration criteria applies to this index.

Case 2: If the index does not fall under Case 1, but the actuary has sufficient credible data about returns for the index in question, then the process has three steps:

a. Perform a model test. The model would first be fitted to the S&P TSX Composite total returns from January 1956 to June 2010. The model outputs are then compared to the calibration criteria for L1 indices. If the model outputs satisfy those criteria, then the form of the model is acceptable and the actuary can proceed to the second step. If not, then the actuary would change the model form.

b. Fit the model to the available data for the index. The model is then used to generate returns.

c. A final test is to review the Sharpe ratio[1] of the model outputs. The Sharpe ratio is to be calculated using the expectation and the standard deviation of the one-year accumulation factor. The Sharpe ratio would not exceed 0.40 with an assumed risk-free rate of 4.00%. If necessary, the fitted parameters for the mean from Step b. would be adjusted downward until this Sharpe criterion is satisfied.

Case 3: If the index does not fall under Case 1 or Case 2, then the criteria to be applied are derived from criteria for the L1 indices with an adjustment for the expected differences in mean returns and volatility.

The criteria for the accumulation factor of the index are:

$$AF(F, p, t) = AF(TSX, p, t) \times \exp(\mu_{Diff} \times t + \sigma_{Diff} \times \Phi^{-1}(p) \times \sqrt{t})$$

where,

$AF(F, p, t)$	is the left tail criterion for index F for the p^{th} percentile at horizon t;
$AF(TSX, p, t)$	is the left tail criterion for L1 indices for the p^{th} percentile at horizon t;
$\Phi^{-1}(p)$	is the inverse cumulative distribution function of the normal distribution;
σ_{TSX}	is the sample standard deviation for the TSX;

[1] The Sharpe ratio of an index is equal to the difference between the expected return of the index and the risk-free rate, divided by the standard deviation of the index.

σ_F is the sample standard deviation for the index;

σ_{Diff} is equal to σ_F - σ_{TSX} , the differential in the standard deviation of the two indices;

μ_{TSX} is the sample mean for the TSX;

μ_F is the mean for the index, calculated using the Sharpe ratio as:

$$\mu_F = r + \sigma_F \times (\mu_{TSX} - r)/\sigma_{TSX};$$

μ_{Diff} is equal to μ_F - μ_{TSX}, the differential in the mean of the two indices; and

r is the risk-free rate established at 4% for the calibration.

The sample volatilities for the TSX and the index would be calculated using the longest common historical period available. The sample mean for the TSX would be calculated using the data from 1956.

At a minimum, the index would be no less volatile than the TSX. If appropriate, the assumed volatility would be adjusted upward to reflect the stated objectives of the index.

The maximum value of 0.40 for the Sharpe ratio in Case 2 has been established by considering historical values for the TSX and the S&P 500. These historical values are presented in appendix C.

The MSCI EAFE is an example of index falling under Case 1, as a well-diversified basket of L1 indices represents about 75% of this index as at March 31, 2011. The TSX and S&P 500 indices also fall under Case 1 as they are L1 indices of large economies. The actuary would apply judgment for determining whether an economy is large enough or whether a basket of L1 indices is diversified enough for falling under Case 1.

The Hang Seng, Topix and NASDAQ are examples of indices that fall under Case 2. They are not broad-based indices of developed non-Asian economies nor are they small capitalization indices, but long time series of data are available for these indices.

The Shanghai Composite is an example of index that falls under Case 3. It is not an L1 index nor an L2 index, and data are only available since 1991.

8 REFERENCES

American Academy of Actuaries. Recommended Approach for Setting Regulatory Risk-Based Capital Requirements for Variable Annuities and Similar Products. Presented by the Life Capital Adequacy Subcommittee to the National Association of Insurance Commissioners' Capital Adequacy Task Force, Boston, MA, June 2005.

Clements, A. E., and S.I. White. "Nonlinear Filtering for Stochastic Volatility Models with Heavy Tails and Leverage". Queensland University of Technology, 2004.

Hardy, M. R. "A Regime-Switching Model of Long-Term Stock Returns". *North American Actuarial Journal* 5 no. 2 (2001): 41–53.

Hardy, M. R. *Investment Guarantees; Modelling And Risk Management For Equity-Linked Life Insurance*. Wiley, NJ, 2003.

MSCI Market Classification Framework. June 2010.

Panneton, C.-M. Mean-Reversion in Equity Models in the Context of Actuarial Provisions for Segregated Fund Investment Guarantees. Segregated Funds Symposium, 2002.

Proceedings. Canadian Institute of Actuaries.

Watanabe, Toshiaki, 1999, "A Non-linear Filtering Approach to Stochastic Volatility Models with an Application to Daily Stock Returns". *Journal of Applied Econometrics* 14 (1999): 101–121.

APPENDIX A: MODELS

To develop calibration criteria, two approaches were used: a bootstrap approach which randomly sampled the historical data set to develop a distribution; and a model approach which used the distribution under the model. The working group felt that of the various models available, each had some advantages and disadvantages. Therefore, the analysis compared the results produced by several different models. The models used were:

- The lognormal model

$$y_t = \mu + \sigma\varepsilon_t$$

  with $\varepsilon_t \sim N(0,1)$

- The regime-switching model with two lognormal regimes

  If in regime 1 $y_t = \mu_1 + \sigma_1\varepsilon_t$

  If in regime 2 $y_t = \mu_2 + \sigma_2\varepsilon_t$

  The transition between regimes follows a Markov process with fixed probabilities:

  - p_{12} probability of going to regime 2 when in regime 1
  - p_{21} probability of going to regime 1 when in regime 2

- The regime-switching model with two regimes incorporating a draw-down factor

  If in regime 1 $y_t = \kappa_1 + \beta_1 d_t + \sigma_1\varepsilon_t$

  If in regime 2 $y_t = \kappa_2 + \beta_2 d_t + \sigma_2\varepsilon_t$

  Where d_t is the draw-down measure at time t, it is given by

$$d_t = Min(0, d_{t-1} + y_t)$$

  The transition between regimes follows a Markov process with fixed probabilities:

  - p_{12} probability of going to regime 2 when in regime 1
  - p_{21} probability of going to regime 1 when in regime 2

- The stochastic volatility model

$$y_t = \mu + \sigma_t\varepsilon_t$$

$$h_t = \omega + \delta h_{t-1} + \upsilon\eta_t$$

$$h_t = ln\sigma_t^2$$

$$e^{½h_t} = \sigma_t$$

$$\begin{bmatrix}\varepsilon\\ \eta\end{bmatrix} \sim N\left(0, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$$

There are several variations of models which bear the generic name of stochastic volatility model. For the model we considered, it is the log-variance which follows an

AR1 process. It is different than the stochastic volatility model used by the AAA (2005), which used a log-volatility process.

Maximum likelihood parameters for these models were determined using monthly historical data. The approach used is well documented in Hardy (2001 and 2003) for the regime switching lognormal model, in Panneton (2002) for the regime switching model with drawdown factor and in Clements and White (2004) for the CSVL model.

APPENDIX B: BOOTSTRAP APPROACH

The bootstrap approach consists of the following algorithm:

1) Select a block size, say of n months;
2) Generate a scenario of returns of length n x k by performing a random draw with replacement of k blocks of size n from the historical dataset, and pasting the blocks in sequence;
3) Repeat step 2) N times, generating a sample of N random scenarios of equity returns of length k x n each; and
4) From the sample of equity returns of step 3), calculate the percentiles needed.

Applying the bootstrapping technique without adjustment would give lower weight to the first and last n observations, where n is the block size. For example, with a block size of 6, the first observation is only present in the first block, while the sixth observation is present in six different blocks. To provide the same weight to every observation, the historical data are extended at both ends with simulated returns so that each observation is present in the same number of blocks. The extension of data is done assuming a simple lognormal model and it is redone each time one of the blocks using data extension is selected. Making sure that the first and last observations are given equal weight is more material the larger the block size and the more these observations are in the tails of the distribution.

APPENDIX C: RESULTS

Left Tail of Distributions

TSX Jan 1956 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical	0.69	0.78	0.87									
Bootstrap 3-month	0.76	0.81	0.88	0.72	0.82	0.95	0.82	0.98	1.21	1.28	1.65	2.20
Bootstrap 6-month	0.73	0.80	0.87	0.69	0.79	0.93	0.78	0.94	1.17	1.19	1.56	2.10
Bootstrap 12-month	0.73	0.82	0.87	0.71	0.81	0.94	0.80	0.96	1.19	1.24	1.60	2.14
LN model	0.80	0.84	0.89	0.78	0.87	0.99	0.91	1.06	1.28	1.47	1.83	2.37
RS2LN model	0.71	0.78	0.87	0.61	0.73	0.89	0.67	0.85	1.10	0.99	1.36	1.94
RS2DD1 model	0.74	0.80	0.88	0.85	0.94	1.05	1.11	1.24	1.42	1.99	2.33	2.81
CSVL model	0.74	0.81	0.88	0.68	0.80	0.95	0.77	0.95	1.20	1.20	1.59	2.17

S&P 500 Jan 1956 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical	0.69	0.80	0.88									
Bootstrap 3-month	0.78	0.84	0.90	0.77	0.87	1.00	0.92	1.08	1.31	1.52	1.91	2.48
Bootstrap 6-month	0.76	0.83	0.89	0.74	0.85	0.98	0.87	1.04	1.27	1.41	1.80	2.37
Bootstrap 12-month	0.76	0.82	0.89	0.72	0.83	0.97	0.85	1.02	1.26	1.37	1.76	2.34
LN model	0.82	0.86	0.90	0.82	0.91	1.03	0.98	1.14	1.35	1.65	2.03	2.59
RS2LN model	0.74	0.81	0.88	0.67	0.79	0.94	0.75	0.93	1.18	1.16	1.55	2.13
RS2DD1 model	0.74	0.81	0.89	0.85	0.95	1.08	1.13	1.29	1.49	2.15	2.53	3.06
CSVL model	0.73	0.80	0.88	0.63	0.76	0.92	0.67	0.86	1.11	0.96	1.31	1.84

FTSE Jan 1956 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical	0.62	0.77	0.87									
Bootstrap 3-month	0.76	0.81	0.88	0.76	0.87	1.02	0.94	1.15	1.44	1.80	2.37	3.25
Bootstrap 6-month	0.72	0.79	0.87	0.71	0.84	1.00	0.88	1.09	1.39	1.65	2.22	3.11
Bootstrap 12-month	0.71	0.80	0.88	0.68	0.81	0.99	0.84	1.06	1.38	1.57	2.16	3.09
LN model	0.78	0.83	0.89	0.79	0.90	1.04	1.00	1.20	1.48	1.94	2.51	3.38
RS2LN model	0.71	0.81	0.90	0.64	0.79	1.00	0.76	1.01	1.36	1.42	2.02	3.00
RS2DD1 model	0.70	0.79	0.89	0.82	0.93	1.07	1.14	1.31	1.55	2.35	2.85	3.58
CSVL model	0.71	0.79	0.88	0.64	0.79	0.98	0.75	1.00	1.33	1.35	1.92	2.82

MSCI EAFE Dec 1969 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical		0.72	0.81									
Bootstrap 3-month	0.75	0.81	0.87	0.69	0.79	0.91	0.75	0.90	1.10	1.07	1.36	1.81
Bootstrap 6-month	0.73	0.79	0.86	0.66	0.76	0.89	0.70	0.86	1.07	0.98	1.28	1.74
Bootstrap 12-month	0.70	0.76	0.83	0.62	0.72	0.86	0.66	0.82	1.04	0.91	1.22	1.68
LN model	0.81	0.85	0.89	0.76	0.85	0.96	0.86	0.99	1.18	1.26	1.55	1.98
RS2LN model	0.71	0.77	0.85	0.59	0.70	0.84	0.61	0.76	0.97	0.79	1.07	1.49
RS2DD1 model	0.73	0.78	0.85	0.74	0.82	0.93	0.89	1.01	1.18	1.39	1.64	2.01
CSVL model	0.77	0.83	0.90	0.75	0.87	1.02	0.90	1.10	1.35	1.54	1.99	2.64

Russell 2000 Jan 1956 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical	0.66	0.75	0.84									
Bootstrap 3-month	0.70	0.76	0.84	0.63	0.75	0.91	0.73	0.93	1.21	1.24	1.72	2.50
Bootstrap 6-month	0.70	0.76	0.84	0.64	0.76	0.92	0.74	0.94	1.23	1.26	1.75	2.54
Bootstrap 12-month	0.69	0.78	0.85	0.66	0.78	0.94	0.79	0.99	1.28	1.39	1.89	2.71
LN model	0.74	0.79	0.86	0.70	0.81	0.96	0.84	1.03	1.31	1.48	1.99	2.78
RS2LN model	0.66	0.74	0.83	0.56	0.69	0.87	0.62	0.82	1.13	1.01	1.47	2.27
RS2DD1 model	0.69	0.75	0.83	0.83	0.94	1.08	1.18	1.36	1.60	2.54	3.05	3.80
CSVL model	0.60	0.70	0.81	0.40	0.55	0.74	0.37	0.55	0.83	0.44	0.73	1.25

S&P 500 Jan 1926 to June 2010	1-year			5-year			10-year			20-year		
	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th	2.5th	5th	10th
Empirical	0.63	0.73	0.86									
Bootstrap 3-month	0.71	0.78	0.86	0.64	0.75	0.90	0.71	0.89	1.13	1.07	1.45	2.02
Bootstrap 6-month	0.69	0.77	0.86	0.62	0.74	0.90	0.69	0.87	1.12	1.05	1.44	2.04
Bootstrap 12-month	0.65	0.76	0.87	0.57	0.70	0.87	0.62	0.80	1.07	0.92	1.29	1.89
LN model	0.75	0.80	0.86	0.69	0.79	0.92	0.77	0.93	1.16	1.19	1.55	2.12
RS2LN model	0.62	0.74	0.87	0.42	0.58	0.81	0.42	0.62	0.93	0.55	0.90	1.52
RS2DD1 model	0.61	0.71	0.84	0.57	0.68	0.84	0.73	0.89	1.11	1.25	1.57	2.06
CSVL model	0.74	0.80	0.88	0.65	0.78	0.94	0.73	0.92	1.19	1.11	1.52	2.14

Mean, volatility and Sharpe ratio

TSX Jan 1956 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	10.60%	10.98%	10.58%	10.09%
Standard Deviation	17.60%	18.71%	17.39%	16.92%
Sharpe ratio	0.375	0.373	0.378	0.360

S&P 500 Jan 1956 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	10.64%	10.82%	10.90%	9.00%
Standard Deviation	16.55%	17.60%	17.50%	16.27%
Sharpe ratio	0.401	0.388	0.394	0.307

FTSE Jan 1956 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	13.93%	14.11%	13.11%	12.76%
Standard Deviation	21.13%	20.95%	19.89%	19.61%
Sharpe ratio	0.470	0.482	0.458	0.447

MSCI EAFE Dec 1969 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	9.16%	9.43%	9.30%	10.60%
Standard Deviation	16.34%	18.21%	17.59%	15.74%
Sharpe ratio	0.316	0.298	0.301	0.419

Russell 2000 Jan 1956 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	14.08%	14.42%	14.02%	10.50%
Standard Deviation	23.87%	24.54%	23.61%	22.92%
Sharpe ratio	0.422	0.425	0.424	0.284

S&P 500 1926 to June 2010	LN model	RS2LN model	RS2DD1 model	CSVL model
Expected return	11.68%	12.01%	11.30%	10.74%
Standard Deviation	21.58%	21.46%	21.44%	19.12%
Sharpe ratio	0.356	0.373	0.340	0.352